

Query Match	100.0%;	Score 2061;	DB 6;	Length 2061;
PD	03-OCT-2002			
PD	22-MAR-2002	WO 2002JP002782		
PR	22-MAR-2001	JP 01P 082227		
PI	KOJI YANAI, NAOMI SUMIDA, MANNABU WATANABE, TATSUKI MORIYA, TAKESHI			
PI	MURAKAMI			
PC	C12N1/5.61, C12N1/15, C12P17/14// (C12N1/15, C12R1:645), (C12P17/14,			
PC	C12R1:645)			
CC	Transformant producing PF1022 substance and process for CC			
CC	producing the same			
CC	and novel biosynthetic gene			
CC	Location/Qualifiers			
CC	key (1) . . (2058) .			
CC	Location/Qualifiers			
CC	1. .2061			
CC	/organism="Streptomyces venezuelae"			
CC	/mol_type="genomic DNA"			
CC	/db_xref="taxon:54571"			
FEATURES	source			
FT	CDS			
ORIGIN				



Best Local Similarity 100.0%; Pred. No. 2, 2e-158;  
Matches 2061; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGGGCAAGGCTTGGATGCAACTAGACCTGCTCAACCAAAACCTTCCAGTACATC 60
Db 1 ATGGGCAAGGCTTGGATGCAACTAGACCTGCTCAACCAAAACCTTCCAGTACATC 60
QY 61 GGGAGGCAACCGGAGCAACCCCGGTGCTGTGCGCCAAACGACGCACTGATGCGGCTG 120
Db 61 GGGAGGCAACCGGAGCAACCCCGGTGCTGTGCGCCAAACGACGCACTGATGCGGCTG 120
QY 121 CCGGTGAGAGACTTTCAGACGATGCTGTGTCCCGGGCCCGGACGCTCCGACCGGAA 180
Db 121 CCGGTGAGAGACTTTCAGACGATGCTGTGTCCCGGGCCCGGACGCTCCGACCGGAA 180
QY 181 CGGGAATTCGGAATCAACCCCGGAGCATGACAGCAACGCGCCCTGCCCCGCTCTCGAGTTC 240
Db 181 CGGGAATTCGGAATCAACCCCGGAGCATGACAGCAACGCGCCCTGCCCCGCTCTCGAGTTC 240
QY 241 TGCTCGGACCAACGAGGCAATGCGCAGCTCTTGGCGGAAACGTCGCGCTCGCCCGGAA 300
Db 241 TGCTCGGACCAACGAGGCAATGCGCAGCTCTTGGCGGAAACGTCGCGCTCGCCCGGAA 300
QY 301 CCCATGCAAGGCGGAGTCTTCGAGGTGCGGCAACCGGAGGACGTCTTCCGAGGCTTC 360
Db 301 CCCATGCAAGGCGGAGTCTTCGAGGTGCGGCAACCGGAGGACGTCTTCCGAGGCTTC 360
QY 361 CCGTCCGCGTTCAACCGCGGTGCGTACACTCCCTGGCCCGACCGACCTTCCCGAGAG 420
Db 361 CCGTCCGCGTTCAACCGCGGTGCGTACACTCCCTGGCCCGACCGACCTTCCCGAGAG 420
QY 421 CTGGAACCCCTCGCTGAGAGCAAGCAAGGAGTGTATGAGGCTTGGCGGCAACCGGAGAG 480
Db 421 CTGGAACCCCTCGCTGAGAGCAAGCAAGGAGTGTATGAGGCTTGGCGGCAACCGGAGAG 480
QY 481 CCGCTGTGGGCGCTCCAGTTCCACCGGAGTCCATGCGGACGCACTTCCGCGGAGATC 540
Db 481 CCGCTGTGGGCGCTCCAGTTCCACCGGAGTCCATGCGGACGCACTTCCGCGGAGATC 540
QY 541 ATGGCAACTTCCGCGGCTGCGCTGCGCCACCAACCGGAGCAAGGCGCAACCGGAGCGAC 600
Db 541 ATGGCAACTTCCGCGGCTGCGCTGCGCCACCAACCGGAGCAAGGCGCAACCGGAGCGAC 600
QY 601 TCCCGTACGAACCTCCAGTGCAGCGCGGTGCAAGTGTGCGGACCGCGGAAGATACGC 660
Db 601 TCCCGTACGAACCTCCAGTGCAGCGCGGTGCAAGTGTGCGGACCGCGGAAGATACGC 660
QY 661 CGGAGCTGCTGCCCGGAGGAGCAACGTTCTGCTGAGACAGACGCTCCGTCTTGAA 720
Db 661 CGGAGCTGCTGCCCGGAGGAGCAACGTTCTGCTGAGACAGACGCTCCGTCTTGAA 720
QY 721 GGGGCTCGGCGCTTCTCTCTCTCGGAGAGACCGCGGCGCGCTCGCGGAGTACTACAC 780
Db 721 GGGGCTCGGCGCTTCTCTCTCTCGGAGAGACCGCGGCGCGCTCGCGGAGTACTACAC 780
QY 781 TACGCGTCCGCAAGGAGTGTCTCGCTCGCGGCTCCGACCGGACCAAGACCGGAGC 840
Db 781 TACGCGTCCGCAAGGAGTGTCTCGCTCGCGGCTCCGACCGGACCAAGACCGGAGC 840
QY 841 CGGCGCCCTTCTTCAACTCTGAGAGACAGCTCGAACGCGACGAGTCCCGTTCGC 900
Db 841 CGGCGCCCTTCTTCAACTCTGAGAGACAGCTCGAACGCGACGAGTCCCGTTCGC 900
QY 901 CCGGAAGTGGCTTTCAGGTTCAACCTGAGGCTACCTCGGCTTACGAGCTGAAG 960
Db 901 CCGGAAGTGGCTTTCAGGTTCAACCTGAGGCTACCTCGGCTTACGAGCTGAAG 960
QY 961 GCGGAGACCAACCGGAGCAACCGGAGCAACCGGAGCAACCGGAGCAACCGGAGCAAC 1020
Db 961 GCGGAGACCAACCGGAGCAACCGGAGCAACCGGAGCAACCGGAGCAACCGGAGCAAC 1020
QY 1021 GCGGAGCGGCGCATCGGCTTGAACCAACGAGAAAGGCTGCTGCTTACCTGAGCCTTGAC 1080
Db 1021 GCGGAGCGGCGCATCGGCTTGAACCAACGAGAAAGGCTGCTGCTTACCTGAGCCTTGAC 1080
```

```
Db 1021 GCGGAGCGGCGCATCGGCTTGAACCAACGAGAAAGGCTGCTGCTTACCTGAGCCTTGAC 1080
QY 1081 CGCGGGGCGACAGACGAGGCGCGCGGCTGCGGAGAGCGGCGGAGACCTTCACC 1140
Db 1081 CGCGGGGCGACAGACGAGGCGCGCGGCTGCGGAGAGCGGCGGAGACCTTCACC 1140
QY 1141 GGCCTGAGCGTCCGCGCCCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1200
Db 1141 GGCCTGAGCGTCCGCGCCCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1200
QY 1201 GCGGCGGCGGCTTCCGCGCCCTTGCCTGCGCGCGCGGCGGAGCAAGAGCGGAGCGGAG 1260
Db 1201 GCGGCGGCGGCTTCCGCGCCCTTGCCTGCGCGCGCGGCGGAGCAAGAGCGGAGCGGAG 1260
QY 1261 CGCATGAGAGAGTCCGCTCAAGAGAGTCCGCAACGCGAGTCCGTAAGAGATTCGCGTAC 1320
Db 1261 CGCATGAGAGAGTCCGCTCAAGAGAGTCCGCAACGCGAGTCCGTAAGATTCGCGTAC 1320
QY 1321 AACATGATCAACGCGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1380
Db 1321 AACATGATCAACGCGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1380
QY 1381 ATCAGCCCGCTCCGTAAGCGGCGCGCTGCTGAGTTCCCGAACTGTCGCTGAGCGCC 1440
Db 1381 ATCAGCCCGCTCCGTAAGCGGCGCGCTGCTGAGTTCCCGAACTGTCGCTGAGCGCC 1440
QY 1441 TCGCCGAGAGGCTTCTCAAGATCGGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1500
Db 1441 TCGCCGAGAGGCTTCTCAAGATCGGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1500
QY 1501 GGGAGCCGCGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1560
Db 1501 GGGAGCCGCGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1560
QY 1561 GGCAGGAGAGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1620
Db 1561 GGCAGGAGAGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1620
QY 1621 AACAGGCTGAGGCGATCGGCTCCGCTCCGCTCAAGTCCCGGCTCTTGGAGGAGGAGCTTAC 1680
Db 1621 AACAGGCTGAGGCGATCGGCTCCGCTCCGCTCAAGTCCCGGCTCTTGGAGGAGGAGCTTAC 1680
QY 1681 GCGCCCGTCAACGAGCTGATGATGACATCGGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1740
Db 1681 GCGCCCGTCAACGAGCTGATGATGACATCGGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1740
QY 1741 GCGGCTGCTGAGCGGCGGCTTCCCGGCGGCTCCATGACCGGCGGCGGAGCGGAGCGGAG 1800
Db 1741 GCGGCTGCTGAGCGGCGGCTTCCCGGCGGCTCCATGACCGGCGGCGGAGCGGAGCGGAG 1800
QY 1801 ACCATGAGATGATGAGCGGCTGAGAGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
Db 1801 ACCATGAGATGATGAGCGGCTGAGAGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
QY 1861 GAGTGTTCGCGCTTCAAGCGGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1920
Db 1861 GAGTGTTCGCGCTTCAAGCGGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1920
QY 1921 GCGGAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980
Db 1921 GCGGAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980
QY 1981 GAGGAGTTCACGAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040
Db 1981 GAGGAGTTCACGAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040
QY 2041 GCGGAGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2061
Db 2041 GCGGAGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2061
```

RESULT 2  
BD093914



LOCUS	BD0931914	2061 bp	DNA	linear	PAT 27-AUG-2002			
DEFINITION	Transformants that produce secondary metabolites modified by a functional group(s) and novel biosynthesis genes.							
ACCESSION	BD0931914							
VERSION	BD0931914.1	GI:22639502						
KEYWORDS	WO 0123542-A/1.							
SOURCE	Streptomyces venezuelae							
ORGANISM	Streptomyces venezuelae							
REFERENCE	Bacteria; Actinobacteridae; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.							
AUTHORS	Yanai, K., Okakura, K., Yasuda, S., Watanabe, M., Miyamoto, K., Mido, N.							
TITLE	Transformants that produce secondary metabolites modified by a functional group(s) and novel biosynthesis genes							
JOURNAL	Patent: WO 0123542-A 1 05-APR-2001;							
	MEIJI SEIWA KAISHA LTD, KOJI YANAI, KAORU OKAKURA, SHOHEI YASUDA, MANABU WATANABE, KOICHI MIYAMOTO, NAOKI MIDO, TAKESHI MURAKAMI							
COMMENT	OS Streptomyces venezuelae							
	PN WO 0123542-N/1							
	PD 05-APR-2001							
	PF 29-SEP-2000 WO 2000JP006783							
	PR 29-SEP-1999 JP 99P 276314							
	PI KOJI YANAI, KAORU OKAKURA, SHOHEI YASUDA, MANABU WATANABE, KOICHI MIYAMOTO, NAOKI MIDO, TAKESHI MURAKAMI							
	PI MIYAMOTO.							
	PI NAOKI MIDO, TAKESHI MURAKAMI							
	PC C12N15/09, C12N5/10, C12P21/02, C07K11/00, C12P21/02, C12R1:645							
FEATURES								
source	FT	Key	Location/Qualifiers					
		1..2061	(1)..(2058).					
		Location/Qualifiers						
		1..2061						
		/organism="Streptomyces venezuelae"						
		/mol_type="genomic DNA"						
		/db_xref="taxon:54571"						
ORIGIN								
Query Match	100.0%;	Score 2061;	DB 6;	Length 2061;				
Best Local Similarity	100.0%;	Pred. No. 2.2e-158;						
Matches 2061;	Conservative	0;	Mismatches	0;	Indels			
				0;	Gaps			
OY	1	ATGGCGCAGCTTGTGATCGACAACTACGACCTGTTCAACCCAACTTTCCAGTACATC	60					
DB	1	ATGGCGCAGCTTGTGATCGACAACTACGACCTGTTCAACCCAACTTTCCAGTACATC	60					
OY	61	GCGCAGGCCAACCGGGCAACCCCGCGTGTGTCGCCAACAAGCCGCACTGTCGGCGTG	120					
DB	61	GCGCAGGCCAACCGGGCAACCCCGCGTGTGTCGCCAACAAGCCGCACTGTCGGCGTG	120					
OY	121	CCCGTCGAGGACTTGACGCGCATGTCGTGCCCGGGGCCCGGAGCCCGGACCCGAGAAC	180					
DB	121	CCCGTCGAGGACTTGACGCGCATGTCGTGCCCGGGGCCCGGAGCCCGGACCCGAGAAC	180					
OY	181	CGGACCTTCGGAATCAGCCGCGCGGGCGATCAACCGACAGCGGCTCCGCTCTGCGCTC	240					
DB	181	CGGACCTTCGGAATCAGCCGCGCGGGCGATCAACCGACAGCGGCTCCGCTCTGCGCTC	240					
OY	241	TGCTTCGGGCACACGAGGCGATGCGCAGCTTTCCGCGGAAACCGTCGCGCTCGCCCGGAA	300					
DB	241	TGCTTCGGGCACACGAGGCGATGCGCAGCTTTCCGCGGAAACCGTCGCGCTCGCCCGGAA	300					
OY	301	CCCATGCAACGCGCGGGGTCTCGAGAGTGGGGCAACCGGCGAGAGAGTCTTCGCGGGGCTC	360					
DB	301	CCCATGCAACGCGCGGGGTCTCGAGAGTGGGGCAACCGGCGAGAGAGTCTTCGCGGGGCTC	360					
OY	361	CCCTGCGGTTTACCGCGCGTGGCTACCACTTCCCTGAGCCGACACCGACCTCCCGACGAG	420					
DB	361	CCCTGCGGTTTACCGCGCGTGGCTACCACTTCCCTGAGCCGACACCGACCTCCCGACGAG	420					
OY	421	CTCGAACCCTTGGCTTGGAGGACGACGCGGGTCTGATAGGGCTTCGGGACACCGGAGAG	480					
DB	421	CTCGAACCCTTGGCTTGGAGGACGACGCGGGTCTGATAGGGCTTCGGGACACCGGAGAG	480					

OY	481	CCGCTGTGGGGCGTTCAGATTTCACCCGGAGTTCATCGGACGACGATTTGGCGGGAGATC	540
Db	481	CCGCTGTGGGGCGTTCAGATTTCACCCGGAGTTCATCGGACGACGATTTGGCGGGAGATC	540
OY	541	ATGGCCAACTTTCGCGACCTTCGCTTCGCTCGCCACCAACCGGGACCGGCGCACCGGGCGGAC	600
Db	541	ATGGCCAACTTTCGCGACCTTCGCTTCGCTTCGCTCGCCACCAACCGGGACCGGCGCACCGGGCGGAC	600
OY	601	TCCTCCGTACGAACTTCACGTGCGCCGCTGTGAAGTCTGTGCCGAGCGCCGAAGAGTACGC	660
Db	601	TCCTCCGTACGAACTTCACGTGCGCCGCTGTGAAGTCTGTGCCGAGCGCCGAAGAGTACGC	660
OY	661	CGGGCGTGCCTCGCCCGGGGAGGGGACACAGTTTCGTGGTAGAGACAGCACTCCGTCTAGAA	720
Db	661	CGGGCGTGCCTCGCCCGGGGAGGGGACACAGTTTCGTGGTAGAGACAGCACTCCGTCTAGAA	720
OY	721	GGCGCCTTCGCGCTTCTCTTTCGAGGAGACCGCGGCGCGCTCGCGGATACCTACAC	780
Db	721	GGCGCCTTCGCGCTTCTCTTTCGAGGAGACCGCGGCGCGCTCGCGGATACCTACAC	780
OY	781	TACCGCGTCCGCGACGCGCGTGTCTCGTCCGCGCTCCGACCGGACACAGACCCGAGCG	840
Db	781	TACCGCGTCCGCGACGCGCGTGTCTCGTCCGCGCTCCGACCGGACACAGACCCGAGCG	840
OY	841	CGGCGCCCTTCTTTCACACTACCTGGAGAGACAGTCCGACCGGACCGGGTCCCGTGGCC	900
Db	841	CGGCGCCCTTCTTTCACACTACCTGGAGAGAGCTCCAGACCGGACCGGGTCCCGTGGCC	900
OY	901	CCCGAATGGCCCTTCAGATTCAACTCGGCTACGTCGGCTACCTCGGCTACCGAGCTGAAG	960
Db	901	CCCGAATGGCCCTTCAGATTCAACTCGGCTACGTCGGCTACCTCGGCTACCGAGCTGAAG	960
OY	961	GCGGAGACCAACCGGCGACCCCGCGCACCGGTCGCCCGACCCCGACCGCGCGTTCCTTTC	1020
Db	961	GCGGAGACCAACCGGCGACCCCGCGCACCGGTCGCCCGACCCCGACCGCGCGTTCCTTTC	1020
OY	1021	GCGGACCGCGCATTCGCTTCGACACCAAGAGAGCTGTCTACTCTGTGCTGGCTTGGAC	1080
Db	1021	GCGGACCGCGCATTCGCTTCGACACCAAGAGAGCTGTCTACTCTGTGCTGGCTTGGAC	1080
OY	1081	CGCGGGGGCCACAGACGAGCGGCGCCCGGCGCTGTGTGGGGAGACGGGCGGAGACCCCTCAC	1140
Db	1081	CGCGGGGGCCACAGACGAGCGGCGCCCGGCGCTGTGTGGGGAGAGCGGCGGAGACCCCTCAC	1140
OY	1141	GGCTTCGCGCTTCGCGCGCCCGGCGGACCGAGCTCCCGCAATGTCTTTCGGGATCCCGAG	1200
Db	1141	GGCTTCGCGCTTCGCGCGCCCGGCGGACCGAGCTCCCGCAATGTCTTTCGGGATCCCGAG	1200
OY	1201	GCGGCGGCGCGGCTTTCGCGCCCTTCGTGGCCGCGCGGACGACAGAGACGCTTACTCAAG	1260
Db	1201	GCGGCGGCGCGGCTTTCGCGCCCTTCGTGGCCGCGCGGACGACAGAGACGCTTACTCAAG	1260
OY	1261	CGCATTCAGACGATGCTTCAGAGAGAGATCCGAGACGAGCGAGTCGATCGATCTGCTGACC	1320
Db	1261	CGCATTCAGACGATGCTTCAGAGAGAGATCCGAGACGAGCGAGTCGATCGATCTGCTGACC	1320
OY	1321	AACATGTGTACCGCGCGGACCGAGGCGACCGGCTTCGCTTACTTCGCGCTTGGCGGCC	1380
Db	1321	AACATGTGTACCGCGCGGACCGAGGCGACCGGCTTCGCTTACTTCGCGCTTGGCGGCC	1380
OY	1381	ATCAGCCCGCTCCGCTACCGCGGCTTCGCTGTGAATTCCTCCGAACCTGTCTGTAGAGGCC	1440
Db	1381	ATCAGCCCGCTCCGCTACCGCGGCTTCGCTGTGAATTCCTCCGAACCTGTCTGTAGAGGCC	1440
OY	1441	TTCGCGCAGAGGCTTCTTCACGATTCGGGCGGACGCGGCGGCTTCGATCTCAAGCCATCAAG	1500
Db	1441	TTCGCGCAGAGGCTTCTTCACGATTCGGGCGGACGCGGCGGCTTCGATCTCAAGCCATCAAG	1500
OY	1501	GGGACCCGCGCCCGGGGCGGACCGCGGAGAGAGACAGCGGCTTCGCGCGGACCTTGGCC	1560
Db	1501	GGGACCCGCGCCCGGGGCGGACCGCGGAGAGAGAGAGAGGCTTCGCGCGGACCTTGGCC	1560
OY	1561	GGCGGAGAGAGACCGGGCCGAGACCTGTATGATCTGTCAACTGTGTTCGCAACGACTTC	1620



D	b	1561	GCCCCGGAGAAAGACCGGGCCGAGAAGACTGATGATCTGTGCACCTTGTCGGAAGCACCCTT	1620
O	y	1621	AAACACGCTCTGCGCATCGGCTCCGCTTCACACGTGCCCGGCTTTTGAGGTGAGAACCTAC	1680
D	b	1621	AACACGCTCTGCGCATCGGCTCCGCTTCACACGTGCCCGGCTTTTGAGGTGAGAACCTAC	1680
O	y	1681	GGGCCCCGTGCACCACTGCTGTGTGCACATCCGGGGAGCGGCTCTGGCACCAACACACC	1740
D	b	1681	GGGCCCCGTGCACCACTGCTGTGTGCACATCCGGGGAGCGGCTCTGGCACCAACACACC	1740
O	y	1741	GGCGCTCGTACGCGCGCGCTTCCC CGGCGGCTCATGACCGGGGCGGCCAAGAAAGCC	1800
D	b	1741	GGCGCTCGTACGCGCGCGCTTCCC CGGCGGCTCATGACCGGGGCGGCCAAGAAAGCC	1800
O	y	1801	ACCATGAGATCATCGACCGCTCTGAGAGAAAGCCCCCGGGGCGTCTATCCGGGGCGCTC	1860
D	b	1801	ACCATGAGATCATCGACCGCTCTGAGAGAAAGCCCCCGGGGCGTCTATCCGGGGCGCTC	1860
O	y	1861	GGATGCTTGCCTTAGCGGGCGCGCCGACCTCAGCATCTGCATCCGACCATGTGCTG	1920
D	b	1861	GGATGCTTGCCTTAGCGGGCGCGCCGACCTCAGCATCTGCATCCGACCATGTGCTG	1920
O	y	1921	GCCGACGGCGAAGCGGAGTTCGGCGGTGGCGGGGGGATCGTGCTCTCCGACCAAGAG	1980
D	b	1921	GCCGACGGCGAAGCGGAGTTCGGCGGTGGCGGGGGGATCGTGCTCTCCGACCAAGAG	1980
O	y	1981	GAGAGATTACCGAGACCGTGTGTAAGGCGCGCCGATGTCACCGCCTTCGACGGAGC	2040
D	b	1981	GAGAGATTACCGAGACCGTGTGTAAGGCGCGCCGATGTCACCGCCTTCGACGGAGC	2040
O	y	2041	GCCGTGCGGGCGGCCGATGA	2061
D	b	2041	GCCGTGCGGGCGGCCGATGA	2061
<hr/>				
RESULT 3				
AB116234				
LOCUS				
DEFINITION				
AB116234 5251 bp DNA linear BCT 03-JUN-2004				
Streptomyces venezuelae pAP, papC, papB, ORFV genes for				
4-amino-4-deoxychorismate synthase, 4-amino-4-deoxyprephenlate				
dehydrogenase, 4-amino-4-deoxychorismate mutase, hypothetical				
protein, complete cds.				
AB116234				
AB116234.1 GI:47846865				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
Streptomyces venezuelae				
Streptomyces venezuelae				
Bacteria; Actinobacteriales; Actinobacteridae; Actinomycetales;				
Streptomycinae; Streptomycetaceae; Streptomyces.				
REFERENCE				
AUTHORS				
Yanai, K., Sunada, N., Okakura, K., Moriya, T., Watanabe, M. and				
Murakami, T.				
TITLE				
Para-position derivatives of fungal antihelminthic cyclodipeptides				
engineered with Streptomyces venezuelae antibiotic biosynthetic				
genes				
JOURNAL				
Nac. Biotechnol. 22 (7), 848-855 (2004)				
PUBMED				
15184904				
REFERENCE				
AUTHORS				
Yanai, K., Sunada, N., Okakura, K., Moriya, T., Watanabe, M. and				
Murakami, T.				
TITLE				
Direct Submission				
Submitted (01-AUG-2003) Koji Yanai, Meiji Seika Kaisha, Ltd.,				
Microbiological Resources and Technology Laboratories, 788 Kayama,				
Odawara-shi, Kanagawa 2500852, Japan				
(E-mail:koji.yanai@meiji.co.jp, Tel:81-465-37-5106,				
Fax:81-465-37-6397)				
FEATURES				
source				
1..5251				
/organism="Streptomyces venezuelae"				

/gene="papa"  
 651..2711  
 /gene="papa"  
 /codon\_start=1  
 /evidence=experimental  
 /transl\_table=11  
 /product="4-amino-4-deoxychorismate synthase"  
 /protein\_id="BAD21140.1"  
 /db\_xref="GI:47846866"  
 /translation="MRTLLIDNYDSFTHTNLFOYIGEAATGCPVVVVPNDADMSRLPVEI  
 FDAIVSVSPGSPDRBDFGISRAATDPSGLPVIGLCEGHOGIADGEGTGTALAPPEP  
 HGRVSEVHTGDEVPFRLPSPTFAVRHSLAATDPELPLASADGCVGTLGRREK  
 PLNGQFIPBESIGDPCGREIMANFRDLALAHRRARRGADSPLYLHRRVDVLDAAE  
 VRRGCLPGEETFTWLDSSSVLEGASRFSFLGDDGPALEYLYTLRAVDGVSVRSDDG  
 TTRRRPFTTLEBQLERRRVPAPELPFEFNIDLVGLEYELAEFTGPAHRSPPH  
 DAAFLFPAARAIALDHQECCLALADRGHDGARMRETAFLTGIAVARAPAEPT  
 AMVGFIPAAAGFGLARARHDKDAYIKRIDECKEIRNGSEYELCTINVTATPAAT  
 ALPLYSALRAISVPYGALEFPEPLVDSLNASPEEPLTIGADGVSEKPIKOTPRGCG  
 AEBEDRLRADLAGREKDAEMNLVDLYRLDINSVCAIGSHVRLFEVEYTAHVQ  
 VSTIRGLRPGTSTACVRAAFPGGSMTGAKRKRTMEIIDLLEGPGVSGALGHQWPA  
 LSGADSLSVIRITVLADGQAEFGVGAIVSLSDQEBEFTETVVKARAMYATLDGSAM  
 AGAR"  
 2708..3676  
 /gene="pabc"  
 2708..3676  
 /gene="pabc"  
 /codon\_start=1  
 /evidence=experimental  
 /transl\_table=11  
 /product="4-amino-4-deoxytryptophanate dehydrogenase"  
 /protein\_id="BAD21141.1"  
 /db\_xref="GI:47846867"  
 /translation="MSGPFRSVVVGSGAGVGFAGLLREAGSRTLVVDLVPPGPRPD  
 ACIVGDTVAPGPPELLAALRDADLVLLVHEEDVALKAVAPYRLRPPALLADTSTVT  
 GMAELIAHAPGVQHVGHVWPFAPGAGTGRPVAAVTRDGPVTALTLVEGGSGRP  
 VRLTAEEHDPRTTAQTALTHAVILSPGLAARLGVDRALAAATPAPHOYIALGAR  
 LGGSPETVGDIGRNSPRAASRRRLAEHLRSFALVGDDPRADAPGADAPGHPGGG  
 DGAENLDGVFGLRLRMGLPELAAQGDHCQELFTLHRTDDEGEKDR"  
 3673..3984  
 /gene="pabp"  
 3673..3984  
 /gene="pabp"  
 /codon\_start=1  
 /evidence=experimental  
 /transl\_table=11  
 /product="4-amino-4-deoxychorismate mutase"  
 /protein\_id="BAD21142.1"  
 /db\_xref="GI:47846868"  
 /translation="MTEONELQRLRAELDALDGLTLDTTTRRIDLVGRTIRAYSRHGV  
 PMMPGPRVSLVKDRAARYAADHGIDSEFLVNLVYDVIITEMCRVEDLVMSRESLSIAEDR  
 R"  
 4065..5243  
 /gene="ORFV"  
 4065..5243  
 /gene="ORFV"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAD21143.1"  
 /db\_xref="GI:47846869"  
 /translation="MHAITTLAALTGRTAQQPYMPDPDAVDVTERELAGLPVLTPEPD  
 EVVLDNGMALVARGALLVOGDCARFHEAVDILVRKVDINIOGLAALIMRAGSGLP  
 AVAALRIAGCYGRSPSPFTEESGAGRKMPSCGDVNVNPEFPAARTEPRLTLRA  
 VTGQIOLVDEIRMSGRPVILERYVTHSELLILPYERDLVREGAHGTSSAHAGMGIG  
 ERTREBGAHYALAAQAVHNPNVGLGTGTVSPEDVALSRLSNBEGVCGRLTFTYRPGA  
 KEGVGLPVPVRAVARHAGAPVWMLCDPMHSGGLAKGKTRLLPEMFAETAFAVTRLR  
 HGWMPAGLHELTPDVTBCVSELDRPPFTDYRSTCDPRLNPEOSADMWTHPLSLIL  
 "



Matches 2061: Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1	ATGCGACGCTTCTGATCGACAACTAGACTCGTTGACCCGACAACTGTTGCTGATCATC	60	
Db	651	ATGCGACGCTTCTGATCGACAACTAGACTCGTTGACCCGACAACTGTTGCTGATCATC	710	
QY	61	GCGGAGGCCACCGGGGACCCCGGTCGTGTCGCCACGACGCGCATGTGCGGGCTG	120	
Db	711	GCGGAGGCCACCGGGGACCCCGGTCGTGTCGCCACGACGCGCATGTGCGGGCTG	770	
QY	121	CCCGTGAAGGACTTCGACGCGATGCTGTCCTCCCGGGCCCGGACGCGCGGAA	180	
Db	771	CCCGTGAAGGACTTCGACGCGATGCTGTCCTCCCGGGCCCGGACGCGCGGAA	830	
QY	181	CGGGAATTGGAATCAGCGCGGGGATGATCGGACGAGCGGCTGCGCGCTCGGGCTC	240	
Db	831	CGGGAATTGGAATCAGCGCGGGGATGATCGGACGAGCGGCTGCGCGCTCGGGCTC	890	
QY	241	TGCTCGGCGCACCGAGGACATCGCCAGCTTTCGGGCGGAACTGCGGCTCGCCGGA	300	
Db	891	TGCTCGGCGCACCGAGGACATCGCCAGCTTTCGGGCGGAACTGCGGCTCGCCGGA	950	
QY	301	CCCATGCAAGCGCGGCTCTCGAGGTGCGGACACCGGCGAGGAGCTTTTCGGGGCTTC	360	
Db	951	CCCATGCAAGCGCGGCTCTCGAGGTGCGGACACCGGCGAGGAGCTTTTCGGGGCTTC	1010	
QY	361	CCCTGCGCGTTTACCGCGCGGCTACCTCCCTGGCGCGCACGACCTTCCCGGACGAG	420	
Db	1011	CCCTGCGCGTTTACCGCGCGGCTACCTCCCTGGCGCGCACGACCTTCCCGGACGAG	1070	
QY	421	CTCGAACCCCTCGCTGAGAGCAGACGAGGCTGTATGAGGCTTGCGGACCGCGAAG	480	
Db	1071	CTCGAACCCCTCGCTGAGAGCAGACGAGGCTGTATGAGGCTTGCGGACCGCGAAG	1130	
QY	481	CGGCTGTGGGCGTTCAGATTCCACCCGAGTTCATCGGACGAGCTTTCGGCGGAGATC	540	
Db	1131	CGGCTGTGGGCGTTCAGATTCCACCCGAGTTCATCGGACGAGCTTTCGGCGGAGATC	1190	
QY	541	ATGGCAACTTCGCGGACCTCGCGCTCGCCACACCGGACAGGGGCCACCGGGGCGAC	600	
Db	1191	ATGGCAACTTCGCGGACCTCGCGCTCGCCACACCGGACAGGGGCCACCGGGGCGAC	1250	
QY	601	TCCCGCTAGAACTCCACGTCGCGCGCTGACGTCGTCGCGGACCGGAAAGATACG	660	
Db	1251	TCCCGCTAGAACTCCACGTCGCGCGCTGACGTCGTCGCGGACCGGAAAGATACG	1310	
QY	661	CGGCGCTGCTGCGCGGACGAGGACACGTTTGTGCTGAGACGAGCTTCGCTCGAA	720	
Db	1311	CGGCGCTGCTGCGCGGACGAGGACACGTTTGTGCTGAGACGAGCTTCGCTCGAA	1370	
QY	721	GCGGCTCGGCTTCTCTTCTCTCGGACGACCGCGGCTCGCTCGGCTGACTTCAC	780	
Db	1371	GCGGCTCGGCTTCTCTTCTCTCGGACGACCGCGGCTCGCTCGGCTGACTTCAC	1430	
QY	781	TACCGGCTGCGGACGAGGCTGCTCGTCCGCGGCTCGGACGAGCAACGACCGGAG	840	
Db	1431	TACCGGCTGCGGACGAGGCTGCTCGTCCGCGGCTCGGACGAGCAACGACCGGAG	1490	
QY	841	CGGCGCTTCTTCACTACCTGAGAGACAGCTCGAACTCGGACGAGGCTCCGCTGCG	900	
Db	1491	CGGCGCTTCTTCACTACCTGAGAGACAGCTCGAACTCGGACGAGGCTCCGCTGCG	1550	
QY	901	CCCGAACTGCTTCACTACCTGAGAGACAGCTCGAACTCGGACGAGGCTCCGCTGCG	960	
Db	1551	CCCGAACTGCTTCACTACCTGAGAGACAGCTCGAACTCGGACGAGGCTCCGCTGCG	1610	
QY	961	GCGGACGAGCGGACCGCGGACCGGCTCCGCGACCGGACGCGGCTTCTCTTC	1020	
Db	1611	GCGGACGAGCGGACCGCGGACCGGCTCCGCGACCGGACGCGGCTTCTCTTC	1670	
QY	1021	GCGGACGAGCGGACCGCGGACCGGCTCCGCGACCGGACGCGGCTTCTCTTC	1080	
Db	1671	GCGGACGAGCGGACCGCGGACCGGCTCCGCGACCGGACGCGGCTTCTCTTC	1730	

QY	1081	CGCGGGGACACGACGAGCGGCGCGGCTGAGTGCAGGAGACGAGCCCTCAC	1140	
Db	1731	CGCGGGGACACGACGAGCGGCGCGGCTGAGTGCAGGAGACGAGCCCTCAC	1790	
QY	1141	GCGCTGCGGCTCGCGCGCGCGGACGAGCCCGGACCATGATCTTCGAGATCCCGAG	1200	
Db	1791	GCGCTGCGGCTCGCGCGCGCGGACGAGCCCGGACCATGATCTTCGAGATCCCGAG	1850	
QY	1201	GCGGCGGCGGCTTCGCGCGCGGCTTCGCGCGGCGGACGAGCAAGAGGCTTACTCAAG	1260	
Db	1851	GCGGCGGCGGCTTCGCGCGCGGCTTCGCGCGGCGGACGAGCAAGAGGCTTACTCAAG	1910	
QY	1261	CGCATGACGAGTGCCTCAAGAGATCGGCAACGAGGATGATGAGATTCGCGTAC	1320	
Db	1911	CGCATGACGAGTGCCTCAAGAGATCGGCAACGAGGATGATGAGATTCGCGTAC	1970	
QY	1321	AACATGATCAGCGCGCGGACCGAGGAGAGCGGCTTCGCTACTCCGCGCTGCGCG	1380	
Db	1971	AACATGATCAGCGCGCGGACCGAGGAGAGCGGCTTCGCTACTCCGCGCTGCGCG	2030	
QY	1381	ATCAGCCCGTCCGTTACGCGCGCTGCTGAGTTCCCGAACTGTGCTGAGCGCC	1440	
Db	2031	ATCAGCCCGTCCGTTACGCGCGCTGCTGAGTTCCCGAACTGTGCTGAGCGCC	2090	
QY	1441	TGCGCCGAGCGGTTCTCAAGATCGGCGCGGACGCGGCGGCTGAGTCCAAAGCCATCAAG	1500	
Db	2091	TGCGCCGAGCGGTTCTCAAGATCGGCGCGGACGCGGCGGCTGAGTCCAAAGCCATCAAG	2150	
QY	1501	GGAACCGCGCGCGGCGGACCGCGAGAGAGAGAGCGGCTCGCGCGGACCTGCGC	1560	
Db	2151	GGAACCGCGCGCGGCGGACCGCGAGAGAGAGAGCGGCTCGCGCGGACCTGCGC	2210	
QY	1561	GCGCGGAGAGAGACCGGCGGAGAACTGATGATGTCGACCTGATCCGAAAGACTTC	1620	
Db	2211	GCGCGGAGAGAGACCGGCGGAGAACTGATGATGTCGACCTGATCCGAAAGACTTC	2270	
QY	1621	AACAGGCTGTGCGCATCGGCTCGGCTCGGCGGCGGCTTCGAGGTGAGAGACTTC	1680	
Db	2271	AACAGGCTGTGCGCATCGGCTCGGCTCGGCGGCGGCTTCGAGGTGAGAGACTTC	2330	
QY	1681	GCGCGGAGAGACCGGCGGAGAACTGATGATGTCGACCTGATCCGAAAGACTTC	1740	
Db	2331	GCGCGGAGAGACCGGCGGAGAACTGATGATGTCGACCTGATCCGAAAGACTTC	2390	
QY	1741	GCGCGCTGCTGACGCGCGCTTTCGCGGCGGCTTCATGACCGGCGGCGGCAAGGCGC	1800	
Db	2391	GCGCGCTGCTGACGCGCGCTTTCGCGGCGGCTTCATGACCGGCGGCGGCAAGGCGC	2450	
QY	1801	ACCATGAGATCATGACGCGCTGAGAGAGAGCGCGGAGGCTTACTTCGCGGCGCTC	1860	
Db	2451	ACCATGAGATCATGACGCGCTGAGAGAGAGCGCGGAGGCTTACTTCGCGGCGCTC	2510	
QY	1861	GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1920	
Db	2511	GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2570	
QY	1921	GCGGACGAGCGGAGGTTGCGGCTGCGGCGGAGATGCTGCTTCGACGAGAG	1980	
Db	2571	GCGGACGAGCGGAGGTTGCGGCTGCGGCGGAGATGCTGCTTCGACGAGAG	2630	
QY	1981	GAGGAGTTCAACGAGACCGTGTAAAGCGCGGCGGATGATGATGATGATGATGATGAT	2040	
Db	2631	GAGGAGTTCAACGAGACCGTGTAAAGCGCGGCGGATGATGATGATGATGATGATGAT	2690	
QY	2041	GCGGTGCGGCGGCGGATGA 2061		
Db	2691	GCGGTGCGGCGGCGGATGA 2711		

RESULT 4  
SVU21728  
LOCUS

SVU21728

2660 bp

DNA

linear

BCT 22-NOV-1996



DEFINITION Streptomyces venezuelae p-aminobenzoic acid synthase (pabAB) gene,  
complete cds.  
ACCESSION U21728  
VERSION U21728.1 GI:1680048  
KEYWORDS  
SOURCE  
ORGANISM Streptomyces venezuelae  
Streptomyces venezuelae  
Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;  
Streptomycinae; Streptomycetaceae; Streptomyces.  
REFERENCE  
AUTHORS 1 (bases 1 to 2660)  
TITLE Brown, M. P., Aidoo, K. A. and Vining, L. C.  
A role for pabAB, a p-aminobenzoate synthase gene of Streptomyces  
venezuelae ISP5230, in chloramphenicol biosynthesis  
Microbiology 142 (Pt 6), 1345-1355 (1996)  
JOURNAL  
MEDLINE 96262706  
PUBMED 8704974  
REFERENCE 2 (bases 1 to 2660)  
AUTHORS Brown, M. P.  
TITLE Direct Submission  
JOURNAL Submitted (26-FEB-1995) Biology, Dalhousie University, Halifax,  
Nova Scotia B3H 4H6, Canada  
COMMENT On Nov 21, 1996 this sequence version replaced gi:1680040.  
FEATURES  
source  
1..2660  
/organism="Streptomyces venezuelae"  
/mol\_type="genomic DNA"  
/strain="ISP5230"  
/db\_xref="taxon:54571"  
<1..507  
/note="ORF1"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAB30311.1"  
/db\_xref="GI:1680049"  
/translation="GSAEYVALPFLGEHEDLRIRKTKMLIRGERSVLPADSTNIS  
PMTYKVAEIVGPDVTFIGMESIGAAAMTIGPLGEPIDRRTDSRLNSNFOQA  
REIVDALPEDVEYVYAMGLEPMWGVMAVDYDESHPAIVDSIDLVRHVODKGGTERL  
HURRLRL"  
651..2660  
/gene="pabAB"  
651..2660  
/gene="pabAB"  
/codon\_start=1  
/transl\_table=11  
/product="p-aminobenzoic acid synthase"  
/protein\_id="AAB30312.1"  
/db\_xref="GI:710438"  
/translation="MRTLIDNYDSFTQNLFOYIGRATGQPPVNPADADWSRLPLEDF  
DAIVSPGSGPDRERDFGISRAITDSGLPVLGVCIGHQIAQLASAPMGRVSEVR  
HTGEDVRFGLPSFTARVYHSLATDLPDELPLAVSDGVVGLSHREKPLMGVQFP  
PESIGDFFREIMANFRDLAHRBARADMDGYELHVRVDVLPDAEVRRAACPA  
EAGTFWLDSSSVLEGASPSPLGDDGPRPLAEYITTVAVAGVVSVRSDGTTTRDAATL  
FYLEQLPBPAPVAPDLPEFNLGTVGLGELKAETTGDPVAPVAPPDADFLAD  
RAIALDHQGCCYLLALDRGHDDGARAWLRETAETLTGLAVRVRPVPVAVGVPE  
AAAGFPLARHDKDASALRNGESYEICLTNMTAPTATLPLYSALRRLSPVGS  
ALPEPELVLSASPERPLTIGADGVESKPIKGTAPRGAPAEDERLRADLAGREK  
RAENLMTIVLVRNDLSVCAIGSVHVPRLFEVGDLPALVHQLVSTTGRRLRPGTSTAC  
VAAFPGSGMTGAPKRPWEIILDRLEBGRGVLPGLMFPALSGAANDLSIVRTIVLA  
DGRAERGVGALVLSDDQEBEFRQTVKARAVTALDSSAVAGAR"

ORIGIN  
Query Match 86.1%; Score 1773.8; DB 1; Length 2660;  
Best Local Similarity 92.8%; Pred. No. 3 2e-135;  
Matches 1913; Conservative 0; Mismatches 97; Indels 51; Gaps 3;  
QY 1 ATGCGACGCTTGTGATCGAACACTAGACTCGTTGACCCCAACCTGTTCCAGTACATC 60  
DB 651 ATGCGACGCTTGTGATCGAACACTAGACTCGTTGACCCCAACCTGTTCCAGTACATC 710  
QY 61 GGGAGGCAACCGGAGCAACCCCGCGTCGTCGCCCAAGCGCGCACTGTCGGCGTG 120  
DB 711 GGGAGGCAACCGGAGCAACCCCGCG---TCGTGCCCAAGCAACCGCACTGTCGGCGTG 767

QY 121 CCGGTGAGAACTTTCGACCGCATCGTGTGTCGCCCGGAGCCCGGACCGGAGAA 180  
DB 768 CCCCTGAGAGACTTTCGACCGCATCGTGTGTCGCCCGGAGCCCGGACCGGAGAA 827  
QY 181 CGGAGCTTTCGAAATGACCGCGCGGAGATACCGACAGCGGCTGCTCGGCGTC 240  
DB 828 CGGAGCTTTCGAGATGACCGCGCGGAGATACCGACAGCGGCTGCTCGGCGTC 887  
QY 241 TGCCTGCGCACACGAGGCAATGCCCGACGCTTTCGAGGAAACCGTCGCTCCCGGAA 300  
DB 888 TGCCTGCGCACACGAGGCAATGCCCGACGCTT-----CGGCGAA 926  
QY 301 CCCATCAGCGCGGAGTCTCCGAGTGGCGGACACCGCGAGAGACGTTCTTCGCGGAGCTC 360  
DB 927 CCCATCAGCGCGGAGTCTCCGAGTGGCGGACACCGCGAGAGACGTTCTTCGCGGAGCTC 986  
QY 361 CCCTGCGGCTTTCACCGCGCGTGCCTTACCACTCCCTGACCGGACCGACTCCCGGACGAG 420  
DB 987 CCCTGCGGCTTTCACCGCGCGTGCCTTACCACTCCCTGACCGGACCGACTCCCGGACGAG 1046  
QY 421 CTCGAACCCCTGCGCTGAGCGGACCAACGAGGATCTCATGAGGCTTGGGTCACCGGAGAG 480  
DB 1047 CTCGAACCCCTGCGCTGAGCGGACCAACGAGGATCTCATGAGGCTTGGGTCACCGGAGAG 1106  
QY 481 CCGCTGTGGGAGGCTCCAGTTCCACCGGAGTCCATCGGAGGCACTTCGCGCGGAGATC 540  
DB 1107 CCGCTGTGGGAGGCTCCAGTTCCACCGGAGTCCATCGGAGGCACTTCGCGCGGAGATC 1166  
QY 541 ATGCGCAACTTTCGCGCACTGCGCTTCCTGCGGACCAACCGGAGCAACCGCGCGCGGCGGAC 600  
DB 1167 ATGCGCAACTTTCGCGCACTGCGCTTCCTGCGGACCAACCGGAGCAACCGCGCGGCGGAC 1226  
QY 601 TCCCGCTAGCACTCCAGTGGCG 660  
DB 1227 TGGGCTAGCACTCCAGTGGCG 1286  
QY 661 CCGCGCTGCTCCCGGCGGAGGAGCAACAGTCTGAGTGGAGCAGCAGTCCGCTCGAA 720  
DB 1287 CCGCGCTGCTCCCGGCGGAGGAGCAACAGTCTGAGTGGAGCAGCAGTCCGCTCGAA 1346  
QY 721 GCGCGCTGCGGCTTCTCTTCTTCTGCGGACGACGCGCGGCGCTGCGCGAGTACTTCACC 780  
DB 1347 GCGCGCTGCGGCTTCTCTTCTTCTTCTGCGGACGACGCGCGGCGCTGCGCGAGTACTTCACC 1406  
QY 781 TACCGGTGCGGAGCGGCTGCTTCCGTCGCGGCTCCGAGCGGACCGGAGCG 840  
DB 1407 TACCGGTGCGGAGCGGCTGCTTCCGTCGCGGCTCCGAGCGGACCGGAGCG 1466  
QY 841 CGGCGGCTTCTTCAACTACCTGAGGAGCAGTTCGAACCGCGGAGCGGCTCCCGTCCG 900  
DB 1467 CGGCGGCTTCTTCAACTACCTGAGGAGCAGTTCGAACCGCGGAGCGGCTCCCGTCCG 1526  
QY 901 CCGGAACCTGCTTCAACTACCTGAGTCACTGCGGCTTCACTCGGCTGAGCTGAAG 960  
DB 1527 CCGGAACCTGCTTCAACTACCTGAGTCACTGCGGCTTCACTCGGCTGAGCTGAAG 1586  
QY 961 GGGAGACCAACCGGAGACCCCGGAGCGGTCGCCGACCCCGAGCGCGGCTTCTTC 1020  
DB 1587 GGGAGACCAACCGGAGACCCCGGAGCGGTCGCCGACCCCGAGCGCGGCTTCTTC 1646  
QY 1021 GCGGACCGGCGCATGCGCTTCGACCAACGAGGAGGCTGCTTACCTGTCGCGCTTCGAC 1080  
DB 1647 GCGGACCGGCGCATGCGCTTCGACCAACGAGGAGGCTGCTTACCTGTCGCGCTTCGAC 1706  
QY 1081 CGCGGAGGCAAGACGAGCGCGCGCTGCGCTGCGGAGACGCGGAGACCTTCACC 1140  
DB 1707 CGCGGAGGCAAGACGAGCGCGCGCTGCGCTGCGGAGACGCGGAGACCTTCACC 1766  
QY 1141 GCGCTGCGCTGCGGCGCGCGCGGCGGAGCGGACCCCGGCACTGCTTGGGATTCGCGAG 1200  
DB 1767 GCGCTGCGCTGCGGCGCGCGCGGCGGAGCGGACCCCGGCACTGCTTGGGATTCGCGAG 1826







misc\_feature 4089..4100  
/note="Region: adenylation domain A4"  
misc\_feature 4350..4370  
/note="Region: adenylation domain A5"  
misc\_feature 4494..4538  
/note="Region: adenylation domain A6"  
misc\_feature 4593..4610  
/note="Region: adenylation domain A7"  
misc\_feature 4647..4706  
/note="Region: adenylation domain A8"  
misc\_feature 4833..4853  
/note="Region: adenylation domain A9"  
misc\_feature 4893..4910  
/note="Region: adenylation domain A10"  
misc\_feature 5043..5075  
/note="Region: thiolation domain"  
misc\_feature 5343..5375  
/note="Region: NADH binding site"

## ORIGIN

Query Match 86.1%; Score 1773.8; DB 1; Length 14159;  
Best Local Similarity 92.8%; Pred. No. 1.6e-135;  
Matches 1913; Conservative 0; Mismatches 97; Indels 51; Gaps 3;

QY 1 ATGCGCAGCGTTTCTGATCGAACACTACGACTCGTTTCAACCAACCTGTTCCAGTACATC 60  
DB 8153 ATCGGACGCGTTCTGATCGAACACTACGACTCGTTTCAACCAACCTGTTCCAGTACATC 8212  
QY 61 GCGAGGCGCACCGGGGCAACCCCGTGTGTCGCCAAGACGCGGACCTGTTGGCGGCTG 120  
DB 8213 GCGAGGCGCACCGGGGCAACCCCGG---TCGTGCCAAGACGCGGACCTGTTGGCGGCTG 8269  
QY 121 CCCGTGAGGACTTTCAGCGGATGTCGTGTCCCGGGCCCGGACGCCGACCGGAGAA 180  
DB 8270 CCCGTGAGGACTTTCAGCGGATGTCGTGTCCCGGGCCCGGACGCCGACCGGAGAA 8329  
QY 181 CCGGACTTTCGGAATCAGCCCGCGGGCGATACCGACAGCGGCTGCCCCGTCTCGCGCTC 240  
DB 8330 CCGGACTTTCGGAATCAGCCCGCGGGCGATACCGACAGCGGCTGCCCCGTCTCGCGCTC 8389  
QY 241 TGCCTCGGCGCACGAGGCGCATGCGCCAGCTCTTGGGGGAAACGTCGCGCTCGGCCGAGAA 300  
DB 8390 TGCCTCGGCGCACGAGGCGCATGCGCCAGCTCTTGGGGGAAACGTCGCGCTCGGCCGAGAA 8428  
QY 301 CCCATGACGCGCGGCTCTCCAGGTGCGGACACCGGCGAGGACGTCTTCCGCGGCTC 360  
DB 8429 CCCATGACGCGCGGCTCTCCAGGTGCGGACACCGGCGAGGACGTCTTCCGCGGCTC 8488  
QY 361 CCCCTCGCGCTTTCACCGCGCGTGCCTACACTCCCTGGCGCGCACCGACTCTCCCGACAG 420  
DB 8489 CCCCTCGCGCTTTCACCGCGCGTGCCTACACTCCCTGGCGCGCACCGACTCTCCCGACAG 8548  
QY 421 CTGGAACCCCTCGCTGGAAGGACGACGCGGGTGTATCGGGGCTGCGGGGCGCGCGAGAG 480  
DB 8549 CTGGAACCCCTCGCTGGAAGGACGACGCGGGTGTATCGGGGCTGCGGGGCGCGCGAGAG 8608  
QY 481 CCGCTGTGGGCGCTTCAGTTCACCCGAGTTCATCGGACGCACTTTCGCGCGGAGATC 540  
DB 8609 CCGCTGTGGGCGCTTCAGTTCACCCGAGTTCATCGGACGCACTTTCGCGCGGAGATC 8668  
QY 541 ATGGCCAACTTTCGCGGACCTCGCGCTTGGCCACCACTGGGACAGGCGGCGGACCGGAC 600  
DB 8669 ATGGCCAACTTTCGCGGACCTCGCGCTTGGCCACCACTGGGACAGGCGGCGGACCGGAC 8728  
QY 601 TCCCGGTAAGAACTTCACGTCGCGCGCTGTCGAGTGTGTCGGACCGCGAAGAGTAAACG 660  
DB 8729 TCCCGGTAAGAACTTCACGTCGCGCGCTGTCGAGTGTGTCGGACCGCGAAGAGTAAACG 8788  
QY 661 CCGCGCTGCTTGCCTCGCGGACGAGGACCACTTTCGCTGGAACAGAGCTCCGTCTCGAA 720  
DB 8789 CCGCGCTGCTTGCCTCGCGGACGAGGACCACTTTCGCTGGAACAGAGCTCCGTCTCGAA 8848  
QY 721 GCGCGCTCGCGCTTTCCTCTTCGCGGACGACCGCGGCGCGCTCGCGGAGTACTTCACC 780

DB 8849 GCGCGCTCGCGCTTTCCTCTTCGCGGACGACCGCGGCGCGCTCGCGGAGTACTTCACC 8908  
QY 781 TACCGGTGCGCGGACGCGGCTGCTTCCTCCGCTCGGACGAGCAACGACCGGAGAG 840  
DB 8909 TACCGGTGCGCGGACGCGGCTGCTTCCTCCGCTCGGACGAGCAACGACCGGAGAG 8968  
QY 841 CCGCGCGCTTTCCTCACTACCTGAGAGAGAGTGAAGAGCGGACGCGGCTCCGCTCGGC 900  
DB 8969 CCGCGCGCTTTCCTCACTACCTGAGAGAGAGTGAAGAGCGGACGCGGCTCCGCTCGGC 9028  
QY 901 CCCGAATCGCCTTTCGAGTTCAACTCGGCTACGTGCTTACCTCGGCTTACGAGCTGAAG 960  
DB 9029 CCCGAATCGCCTTTCGAGTTCAACTCGGCTACGTGCTTACCTCGGCTTACGAGCTGAAG 9088  
QY 961 GCGGAGACCAACGCGGACCGCGCGGACCGGCTCCCGACCGGACCGGCGGCTTTCCTTC 1020  
DB 9089 GCGGAGACCAACGCGGACCGCGCGGACCGGCTCCCGACCGGACCGGCGGCTTTCCTTC 9148  
QY 1021 GCGGACCGGCGCATCGCCTTCGACACACGAGAAAGCTGCTTACCTGCTGCGCTCGAC 1080  
DB 9149 GCGGACCGGCGCATCGCCTTCGACACACGAGAAAGCTGCTTACCTGCTGCGCTCGAC 9208  
QY 1081 CCGCGGCGGCGACGACGCGCGCGCTGCTGCTGCGGAGACGCGCGGACCTTCACC 1140  
DB 9209 CCGCGGCGGCGACGACGCGCGCGCTGCTGCTGCGGAGACGCGCGGACCTTCACC 9268  
QY 1141 GCGCTGCGGCTCGGCGCGCGCGCGGCGGACCGCGGACCTTCGAGTCTTGGGATCCCGAG 1200  
DB 9269 GCGCTGCGGCTCGGCGCGCGCGCGGCGGACCGCGGACCTTCGAGTCTTGGGATCCCGAG 9328  
QY 1201 GCGGCGGCGGCTTTCGCGCGCGCGCGGCGGCGGCGGACGAGACGCTTACCTCAAG 1260  
DB 9329 GCGGCGGCGGCTTTCGCGCGCGCGCGGCGGCGGCGGACGAGACGCTTACCTCAAG 9380  
QY 1261 CGGATGACGAGTGCCTCAAGAGATCCGCAACGCGGAGTCTGAGAGTCTGCTGACC 1320  
DB 9381 -----CGGCGCTCCGCAACGCGGAGTCTGAGAGTCTGCTGACC 9421  
QY 1321 AACATGTACACCGCGCGGCGGACGAGGCGGACGCGGCTGCGCTTACCTCGCGGCTGCGGCG 1380  
DB 9422 AACATGTACACCGCGCGGCGGACGAGGCGGACGCGGCTGCGCTTACCTCGCGGCTGCGGCG 9481  
QY 1381 ATCAGCCCGCTCCGTAACGCGCGCGCTTCGAGTTTCCCGAACTGTGCTGAGCGGC 1440  
DB 9482 ATCAGCCCGCTCCGTAACGCGCGCGCTTCGAGTTTCCCGAACTGTGCTGAGCGGC 9541  
QY 1441 TCGCCGGAAGCGGTTCTTCAAGATCGGCGCGGACGCGGCTGAGTCCAAAGCCATCAAG 1500  
DB 9542 TCGCCGGAAGCGGTTCTTCAAGATCGGCGCGGACGCGGCTGAGTCCAAAGCCATCAAG 9601  
QY 1501 GGGACCGCGCGCGGCGGCGGCGGCGGAGGAGGAGCGGCGCTCGCGCGGACCGGCTGCGG 1560  
DB 9602 GGGACCGCGCGCGGCGGCGGCGGCGGAGGAGGAGCGGCGCTCGCGCGGACCGGCTGCGG 9661  
QY 1561 GGC CGGAGAGAGACCGGCGCGGACGAGAACTGATGATGTCGACTGATCCGACGACCTTC 1620  
DB 9662 GGC CGGAGAGAGACCGGCGCGGACGAGAACTGATGATGTCGACTGATCCGACGACCTTC 9721  
QY 1621 AACAGGTTGCGGCGGATCGGCTTCCTTCCAGCGTGCCTCGGCTTTCGAGGTGAGACCTTAC 1680  
DB 9722 AACAGGTTGCGGCGGATCGGCTTCCTTCCAGCGTGCCTCGGCTTTCGAGGTGAGACCTTAC 9781  
QY 1681 GCGCGGTAAGACAGGTCGTCGACGATCGGAGGAGCGGCTCGGCGGCGGACCGGACGAC 1740  
DB 9782 GCGCGGTAAGACAGGTCGTCGACGATCGGAGGAGCGGCTCGGCGGCGGACCGGACGAC 9841  
QY 1741 GCGCGCTGCGTAAGCGCGCGCTTTCGCGGCGGCTCCATGACCGGCGCGGCGGACCGGAC 1800  
DB 9842 GCGCGCTGCGTAAGCGCGCGCTTTCGCGGCGGCTCCATGACCGGCGCGGCGGACCGGAC 9901  
QY 1801 ACCATGAGATATGACCGGCTTGAAGAGGCGCGCGGCGGCTTACCTCGGCGGCTTC 1860



Db	9902	CCCATGGAGATCATCGACCGCTGGAGAAAGCCCGGGGCGTCTTACCCGGGGGCGTC	9961
Qy	1861	GGATGTTTGGCCCTCAGCGCGCGCCGACCTACAGATCTGTATCCGCACCATCTGTCTG	1920
Db	9962	GGATGTTTGGCCCTCAGCGCGCGCCGACCTACAGATCTGTATCCGCACCATCTGTCTG	10021
Qy	1921	GCCGACGGCCAGCGGAGTTTCGGCGTTCGGCGGGGCGATCTGTCTCCCTCCGACGAGAG	1980
Db	10022	GCCGACGGCCGGGCGGAGTTTCGGCGTTCGGCGGGGCGATCTGTCTCCCTCCGACGAGAG	10081
Qy	1981	GAGGATTTCAACCGAGACCGTGTAAAGGCCCGGCCATGTCTACCGGCTTCGACGGCAGC	2040
Db	10082	GAGGATTTCAAGGCGAGACCGTGTAAAGGCCCGGCCATGTCTACCGGCTTCGACGGCAGC	10141
Qy	2041	GCCGTGGCGGGGCGCCGATCA	2061
Db	10142	GCACTGGCGGGGCGCCGATCA	10162

RESULT 6  
AR198361

LOCUS	AR198361	2220 bp	DNA	linear	PAT 20-APR-2002
DEFINITION	Sequence	14	from patent	US 6352839.	
ACCESSION	AR198361				
VERSION	AR198361.1	GI:20246210			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				

REFERENCE  
1 (bases 1 to 2220)  
Blanc, V., Thibaut, D., Bamas-Jacques, N., Blanche, F., Crouzet, J.,

```

TITLE      Streptogramins for preparing same by mutasynthetis
JOURNAL    Patent: US 6352839-A 14 05-MAR-2002;
FEATURES   location/Qualifiers
SOURCE     1. .2220

```

```

/mol_type="unassigned DNA"
ORIGIN

```

Query Match	39.8%;	Score	820;	DB	6;	length	2220;	
Best Local Similarity	64.4%;	Pred.	No.	4.4e-58;				
Matches 1356;	Conservative	0;	Mismatches	655;	Indels	93;	Gaps	5;

QY	2	TGCGACGCTTCTGATCGAACAACTGACGCTGTTACCCACAACCTGTTCAGTACATCG	61
Db	71	TGCGAACCTCTCTGATCGAACAACTGACGCTGTTACCTTACCACTCTTCCAGATGCTGG	130
QY	62	GCGAGGCCACCGGGGCMACCCCCGTGTGTGTCGCCACGACGCGGACTGTGTGGGGCTCG	121
Db	131	CCGAGGTGTGMAAGGGGGCGGCTCCGGTGTGTGTCCGMAAGACGACCCGGCACCTGGCAGG	190
QY	122	CCGTTCAG-----GACTTGCACGGGATGTGTGTGTCCCGGGCCCGGGCAGCCCGAAC	175
Db	191	CCCTGTGGCCCGGGGCGACTTTCGACAACTGTGTGTCTTCACCCGGGCCCGGCACCCCGCCA	250
QY	176	GGGAACGGGGAATTGGGAATCAGCCGCGGGGGGATATCAGGACAGGGCGCTTGGCCGTCTCG	235
Db	251	CCGACACCGGACTTGGGCTCTAGCCGCGGGGTGATATCCGAATGGGACTCTGGCTGTCTCG	310
QY	236	GCGTCTGCTTCGGCCACCGAGGACATTCGCCAGCTCTTTCGGGGGAAACCTTCGGGCTTCGCC	295
Db	311	GGGTGTGCTGGGGCCACCGAGGCCCTGTGTCTGTCTGCGGGGCGCGCGTCTGTCCACGAC	370
QY	296	CGGAACCATTCAGAGGCGCGGGTCTTCCGAGGTGCGGCACACCGGCGAGGAGCTCTTTCGGGG	355
Db	371	CCGAACCTTTCACGGGCGGCACAGCGCATTCGCGCACGAGCGGGGAGGGCTCTGTTCGCGA	430
QY	356	GCCTTCCCTCGCCGTTTCAACCGCGCGGTACACTCCCTGGCGGCCACCGACCTCCCGCG	415
Db	431	ACATTCCTTCCCGCTGACCGCTGTGTGTGCGCTACACTGTCTTACCGTCTCGGCAACTGCGCG	490

OY	416	ACGAGCTCGAACCCTCGCCCTGGAGGACGACGACGAGGCTGTCATGGGCGCTGGGACACCGG	475
Db	491	CCGACCTCGCGCCACCGCCCAACCGCGACGGCAGCTAATGGCCCTTCGCCACCGC	550
OY	476	AGAACCGCTCTGGGGGCGTCCAGTTTCCACCCGAGTCATCGAGCAGCACTTCGCGGGG	535
Db	551	ACCTGCCCCGTTGGGCGTGGAGTTTCCACCCCGAATGCATGACGACGACACCGCCAC	610
OY	536	AGATCATGAGCCAACTTCGCGACCT-----	560
Db	611	GGATGCTCGCAACTTCGCGACCTGTCTCTGCGCGCGGCGGACCGACCGCCCCGAC	670
OY	561	-----CGCCCTCGGCCAACCAACCGGACAGGCGGCCACCGGGCGCACTTCC-	604
Db	671	CCGAACGATACCCGCAACCCGACCCGCGCCCCCGCCCCCGCCCGCACCGGACCGCCCG	730
OY	605	-----CGTACGAACTCAAGTGGCGCGGCTCGACGCTGCTGCGGACG	646
Db	731	CGTCCGCGCGCGGTGGGGAGTACCGGCTGCATGTGCGGAGGTGCGCTGCGTCCGACG	790
OY	647	CCGAAGAGTACGCGCGGCTGCTCTGCGCGCGAGGAGCACCACTTCTGTGCTGACAG	706
Db	791	CGACGCGCGGTTTACCGGCTGTGTGCGCGACGCCCGGCGCGGTTCTGGCTGCACAG	850
OY	707	GCTCCGTTCTTGAAAGGCGCTTGGCGCTTCTTCTTCTTGCGGACGACCGCGGCCGCTCG	766
Db	851	GCCCGCTGACCCGGGCGCTCGCCCGCTTTCATCTTCTGCGGCGCCCGCCCGCGCTCG	910
OY	767	CCGAGTACCTGACCTACCGGCTGCGCGGCGCGGCGGCTGCTCGGTCGCGGAGCTCCGACG	826
Db	911	GCGAACAGATACCTTACGACCTGCGCGACCGGCGCGGCGCGCTCAGAGCAGCTTCAGCG	970
OY	827	CCAGCACCGGACCGCGGCGCCCTTCTTTCATCTACCTTGAGAGACAGCTCGAACGCCAC	886
Db	971	GCGAGACCGCGCGGCGCGGACCGCTTTCGACCACTTGAAACAGAACTGG---CGGCC	1022
OY	887	GGGTCCCCGTCGCCCGGAACTGCCCTTTCGAGTTTCACTCGGCTACGTGCGTACCTCG	946
Db	1028	GCGCGCTGCGCGCACCGGCGCTGCGCTTCGAGTTTCACTTGGCGTACCTGCGCTACCTCG	1087
OY	947	GCTACGAGCTGAAGCGGAGACCAACCGGCGACCCCGGCGACCGGTCGCCGACCCCGACG	1000
Db	1088	GCTACGAGACCAAGGCGGACAGCGGCGCGGACGCGCCACCGCGGCAACTGCTCCGACG	1144
OY	1007	CCGCGTTCTTTCGCGGACCGCGCATGCGCTTCGACCAACAGAGAAAGTTCCTTAC	1066
Db	1148	GCGCTTCATGTTTGCGGACCGGATGTCTGCGCTTCGACCAAGAACAGAGGGGGGCTGCG	1207
OY	1067	TGCTGCGCTTCGACCGCGGGGCGACGACGAGCGCGCGCTGCTGCTGCGGAGACGCG	1122
Db	1208	TCTGTGACATGAGCAGCACCGGACGCGCGCGCAACCGCACCGCGC-----G	1252
OY	1127	CCGAGACCTTACCGGCGCTGCGCTTCGCGGCGCGCGCGGACCGACCTCCCGCATGTCT	1188
Db	1253	CCGAACGCTGAGCTTACCGAGCGCGCGCGGACCTTCGCGACCAACCGCGCCCGCGCGCT	1312
OY	1187	TGCGGATTCGCCGAGCGCGGCGCGGCTTTCGAGCCCTTCGACCGCGCGCGCGCACGACAAG	1246
Db	1313	TCAACCTGCTGCGCGGACGACCAACTGCCCGCGCTGAGAGTCTCACTACCGGCACAGCTCG	1377
OY	1247	ACGCTTACTTCAAGCGGATTCGACGAGTTCCTCAAGAGAGATTCGCGACCGGCGAGTCTGACG	1306
Db	1373	CCGCTTACCGGGAACCTGTGAGAGAAATGCGCGCGCTGATCAACGACGGCAGACCTTACG	1433
OY	1307	AGATTCGCGTGAACCAATGTGACCGCGCGGACCGAGCGGACGCGCCCTGCGCGCTTACT	1366
Db	1433	AGGTGTCTTGACGAACATGCTTCGGGTGCGCGCGCGGATTCGACCCGCTCACCGCTTAC	1492
OY	1367	CCGCGCTGCGCGCATCAGCGCGGCTCCGTAACGCGCGCTTCGAGTTCGCCGAACCTGT	1428
Db	1493	GCGCGCTGCGGACCGTACGCGCCCGCGCTTACGCGCGCTTACTCTGACAGTTTCCCGGGGCA	1552
OY	1427	CGGTGTAAGCGCTCGCCCGAGCGATTCTTCAAGATCGGCGCGGACGCGCGCTCGAGT	1486



Db	1553	CCGCGTCACTCTCTCTCCGAAAGGTTCCCTGCGCATCGCGCGAGCGTTGGCGGAGT	1612
Qy	1487	CCAAAGCCATCAAGAGGAGCCCGCGCGGCGCACCGGGAGAGAGAGCGGCTCC	1546
Db	1613	CCAAACCCATCAAGAGGAGCCCGCGCGGCGCACCGGGAGAGAGAGCGGCTCA	1672
Qy	1547	GCGCGCAGCTTGCGCGCGCGGAGAGAGCCCGGCGGAACTGATGATGTCGACCTG	1606
Db	1673	AGGCTCTCTCTGCGCGCGCGAGAGAGCCCGAGCGGAACTGATGATGTCGACCTG	1732
Qy	1607	TCCGCAACGACCTTCAAGAGCTCTGCGCGGATCGGCTCCGTCACGTGCCCCGCTTTCG	1666
Db	1733	TCCGCAACGACCTTCTGCGCAAGTCTGCGACATCGGCTCCGTCACGTACCGGCGCTTTCG	1792
Qy	1667	AGGTGAGACCTTCCGCGCGCGGTGACCGAGTGTGACCATCCCGGCGAGCGGCTGCGC	1726
Db	1793	AGGTGAGACCTTCCGCGCGCGGTGACCGAGTGTGACCGAGTGTGCGCGCGCTGCGC	1852
Qy	1727	CCGCGCACGACCGCGCGCGCTGCGTACGCGCGCGCTTCCCGCGCGCTTCATGACCGCG	1786
Db	1853	CCGCGCACGACCGCGCGCGCTGCGTACGCGCGCGCTTCCCGCGCGCTTCATGACCGCG	1912
Qy	1787	CGCCCAAGAACCGACCATGAGATCATCGACCGCTGAGAGAAAGCGCCCGCGCGCTT	1846
Db	1913	CGCCCAAGAACCGACCATGAGATCATCGACCGCTGAGAGAAAGCGCGCGCGCTT	1972
Qy	1847	ACTCGCGCGCGCTCGAGTGGTTCGCTCGACGCGCGCGCGCGCTTCAGCATGTCATCC	1906
Db	1973	ACTCGCGCGCGCTCGAGTGGTTCGCTCGACGCGCGCGCGCGCTTCAGCATGTCATCC	2032
Qy	1907	GCAACCATGTCGTGCGCGCGCGCGAGCGGAGTTCGCGCGCGCGCGCATGTCGCC	1966
Db	2033	GCAACCATGTCGTGCGCGCGCGCGAGCGGAGTTCGCGCGCGCGCGCATGTCGCC	2092
Qy	1967	TCTCCGACCGAGAGAGAGATTCACCGAGACCGTGTGTAAGCGCGCGCATGTCACCG	2026
Db	2093	TCTCCGACCGCGAGAGAGATTCACCGAGACCGTGTGTAAGCGCGCGCATGTCACCG	2152
Qy	2027	CCCT 2030	
Db	2153	CCCT 2156	
RESULT 7			
SPU60417			
LOCUS			
DEFINITION	SPU60417 4740 bp DNA linear BCT 07-MAR-1997		
ACCESSION	Streptomyces pristinaespiralis 4-dimethylamino-L-phenylalanine		
VERSION	precursor biosynthesis (papA, papC, papB, papM) genes, complete		
KEYWORDS	cds.		
SOURCE	U60417 GI:1575335		
ORGANISM	U60417.1 GI:1575335		
REFERENCE	Streptomyces pristinaespiralis		
AUTHORS	Streptomyces pristinaespiralis		
	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
	Streptomycinae; Streptomycetaceae; Streptomyces.		
	1 (bases 1 to 4740)		
	Blanc, V., Gil, P., Bamas-Jacques, N., Lorenzon, S., Zagorec, M.,		
	Schleuniger, J., Bisch, D., Blanche, F., Debussche, L., Crouzet, J. and		
	Thibaut, D.		
	Identification and analysis of genes from Streptomyces		
	pristinaespiralis encoding enzymes involved in the biosynthesis of		
	the 4-dimethylamino-L-phenylalanine precursor of pristinamycin I		
	Mol. Microbiol. 23 (2), 191-202 (1997)		
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
AUTHORS			
TITLE			
JOURNAL			
	2 (bases 1 to 4740)		
	Blanc, V., Gil, P., Bamas-Jacques, N., Lorenzon, S., Schleuniger, J.,		
	Bisch, D., Blanche, F., Debussche, L., Crouzet, J. and Thibaut, D.		
	Direct Submission		
	Submitted (11-JUN-1996) Recherche Pharmaceutique, Rhone-Poulenc		
	Rorer, 13 quat Jules Guesde, B. P. 14, Vitry sur Seine cedex 94403,		

FEATURES	France
source	Location/Qualifiers
gene	1..4740
CDS	/organism="Streptomyces pristinaespiralis"
	/mol_type="genomic DNA"
	/db_xref="taxon:38500"
	68..2227
	/gene="papA"
	68..2227
	/gene="papA"
	/function="p-aminobenzoate synthase"
	/codon_start=1
	/transl_table=11
	/product="PAPA"
	/protein_id="AAC44866.1"
	/db_xref="GI:1575336"
	/translation="MKTVTLTIIDYDSFTYNLFQMLAEVNGAALPVNRDDTRFQWA
	LAPGDFDNVVSPPGHPATDIDLSRVSITFEMDLPLGVCLHQALCLAGAAVH
	AEPFPGRTSDIRHDGQGLFANI PSLTVVRVHSLTVRLPADLRAHTAHTAGOLAVH
	ARHLPRFGVGHPESSISEGHRLMADRSLRAAGRPHTERIIPAPAPAPAP
	APAPASAPVGEYRLHVREVAQVPDADAFTLAPADPARFWLDSRVDPGLAREFTL
	GNAPGLGQITTYDVADRVRVKGDSGGETRRPFTLPHLEHRLARALPATGLPEEP
	NUGVGYLGEYTKAUSGGEDAHGELPDGAEMFADRMALDHEQGAMALUSTTRP
	ATAPAEERWLTDAARTLATAPRPFTLLPDQPLADVHNSLPRYRELVEBCRL
	INDGETVEVCLTNMLRVGRIDPLTAYRALRTVSPAPAYLQFPGATVLSGSPERFL
	RIGADGMASRKPDKGTRPGAGPADDAVKAASLAEKDSSNNLMIVIDLVRLDGVC
	DIGSVHVPGLFEVETIYATVHOLVSTRGLADVSRPVRAPRAAPFGGSMVGA PKYRTM
	OPRIDPLEKGPBRVYGALCYFALSGAADLSYTRITVATEBAATTIGGAAVVALSDPD
	DEVREMLKAQTTTLALRQAHAGATSRDELHAGSR"
	complement (2486..3376)
gene	/gene="papC"
CDS	complement (2486..3376)
	/gene="papC"
	/function="dehydrogenase"
	/codon_start=1
	/transl_table=11
	/product="PAPC"
	/protein_id="AAC44867.1"
	/db_xref="GI:1575337"
	/translation="MRGGSVFRGCYVNGAGAGVGRMFSHWLVRGVAATWLDVAGAG
	ADGVRVACDVRPPGEAVNALADADVVLAVPEFVAWEAVEVLGVRPGAVLADTL
	SVKSRILAGRLREAPGLQAVGINTPMFASLGIQGRVAAVAVTDPGVALVELVAGW
	GARVEMPARRHDELTLAQQAATHAVALFGLGLSELVDVGLALDRSDAPPPLAMIAL
	LARIAGTPEVYFDIOANPGAPARQALGRGLVRLQGAVERGDEBTFPALPAELRGV
	IGEHGAELERLCARMETALH"
	3413..3802
	/gene="papB"
	3413..3802
	/gene="papB"
	/function="mutase"
	/codon_start=1
	/transl_table=11
	/product="PAPB"
	/protein_id="AAC44868.1"
	/db_xref="GI:1575338"
	/translation="MTPPAIPAPAPATGPAATDPLDRLARLDAADAALDAVTRTL
	DICLRIGEKRLHOVPMOPHRIAQVHNAARVADHDGIDPFLATIVDTITTECRLL
	EBEWLASGAPVPTPVHASARCAVS"
	3799..4677
	/gene="papM"
	3799..4677
	/gene="papM"
	/function="N-methylase"
	/codon_start=1
	/transl_table=11
	/product="PAPM"
	/protein_id="AAC44869.1"
	/db_xref="GI:1575339"
	/translation="WTTAAPTLAQQLDRAATGOLNAGITADAAADPTLLAAHACOVA
	PDODTICLAGVPPRFPMHYVRRRLTRPRAERL VGARVMGHRFDLAPGVFKPETER
	ITRDAIARLEALVRGTTAPLVLDICAPGTMVAATVLAHVARVVLGIELSQAARAA
	RNMGAGTARIVOGARDAPFELSGTVLVTNPPYIVIGLRTSAPVELVHDPPLALW
	AGEEGIGMIRAMERTARLLAPGVLLLEHSGYQLASVPALFLAFRTWSHASSRPTCN







RESULT 8  
STMPABA 4607 bp DNA linear BCT 31-AUG-1993  
LOCUS Streptomyces griseus pab gene fragment.  
DEFINITION M3058  
ACCESSION M3058 GI:153396  
VERSION M3058.1 GI:153396  
KEYWORDS pab gene.  
SOURCE Streptomyces griseus  
ORGANISM Streptomyces griseus  
Bacteria; Actinobacteria; Actinobacteriales; Streptomyces.  
REFERENCE 1 (bases 1 to 4607)  
Crisado, L.M., Martin, J.F. and Gil, J.A.  
The pab gene of Streptomyces griseus, encoding p-aminobenzoic acid synthase, is located between genes possibly involved in candididin biosynthesis  
JOURNAL Gene 126 (1), 135-139 (1993)  
MEDLINE 93231527  
PUBMED 8472954  
COMMENT Original  
FEATURES  
source  
1..4607  
/organism="Streptomyces griseus"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1911"  
/issue\_1fb="IMR03570"  
88..1173  
/note="thioesterase-II-S-acyl fatty acid synthase homolog"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AA072110.1"  
/db\_xref="GI:388262"  
/translation="MRVTDSQCVCAGOCVLNAPVFPDODDDVVVLRADPTSGTT  
RBRARATCARRRSSRRREPRGACDRRCVPYRPGRPADPPAPMDPRQNR  
DSRYTADPTAAMWLRHYPAEDAVLYICFPHAGSAGTYHVSARFAPGAEVSLQ  
YPRQDRKRPVCDLGLTADLITTEQLPDERPVPFGHSGAALAFETAMLEQK  
AGRPYIASGRGSPSTTRAEVHTRDDGIVAMKRLNGTAGVLDDEBILMLPAL  
RGDYRAIEVYTCPPDRRLACGLTTLTGEDDPLTVEABERMDHTGTPRLRYFTGSH  
FPLTQHLDAVNTIFAQALHPDRAAPAA"  
1370..3541  
/gene="pabAB"  
1370..3541  
/gene="pabAB"  
/codon\_start=1  
/transl\_table=11  
/product="p-aminobenzoic acid synthase"  
/protein\_id="AA072111.1"  
/db\_xref="GI:388263"  
/translation="MRTLLVDNYDSFTYNYLPHYLSRANGREBEVIRNDPAPMRPGILD  
APFNVLSPGPGPHRPADFGCLIAEGRILVIGVCGHOGMALAHGARGARPEP  
RGRTSAVHDGTGLFEGGLPOLLEVRYSLATVELPELLEATMASEDDVLMALRRT  
LPLMGVQFPHESIGTODGHRLLANFRDLTBERGRTPHGGRAGGCTLPAPARERTAT  
TGRPRRLVIAKSLPTRMDAEVAFDSLFTGHPFVLDSRPEGELGQLSMGDASGP  
LARTACADVHAGTVVRADGASVSAFLTWLENDLAGLRTVEPLPAPAFMGWGC  
LGELEACDGDGAHRSDDPDAVLVADADALVLDHRTTYLLALVEDDAEABAMWL  
AAASATLDVAGREPPCPPEAPVCTGPVELRHDRDGYLKLIDVCOEIAAGETVEVC  
LNNMAEDTDLFPMAVYRALRVSPAPFAFLDPGMAVLSSPFRFLRIDHGMES  
KPLKGRPRGATPOEDALVRALATCEKRAENIMVLDLVRHDLGRCAVGSVADPV  
FOVETATYHQLVSTYTTARLREDSFVAALVRAPFGSMTGPKRTWQITDLREGP  
RGVYSGAIGFSLTGAVDLSIVRTIVLSGGRILRGVGAVALALSDPADEFETAVKA  
APLLRLDTPAFRRERPGKDLDEEPDDGTDAGAPKDLVLPG"  
3685..>4607  
/note="p-comparate: CoA ligase homolog"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AA072112.1"  
/db\_xref="GI:388264"

ORIGIN  
Query Match 37.5%; Score 773.4; DB 1; Length 4607;  
Best Local Similarity 63.2%; Pred. No. 1.9e-54;  
Matches 1341; Conservative 0; Mismatches 686; Indels 96; Gaps 6;  
/translation="MVPVHAHDVYDTPSTGRTLDGLTLPVFADAVRHGDVAVL  
DEGVALTWGAHRTAVDALRGLOESGVSDVAVLHLPMSWEYLTHLAAAVGAYTM  
PVHOGNAPSDVALLERVPAPAVULATPORGCGTGPALREVPRLRAYVYTGDA  
GGSTEVYEMLRMSGEDLPVYVPPDSFLLPSSGTTSAAPKCTLSHEGLTNR  
AAETEDADAVAGTILTACPLTCHFGIQSAVSALFPARGVLLSGMDVGRFLELRER  
PSVVAVAVPASCITWSPGCAKRTRTAPASARPDP"  
QY 1 ATGCGACCGCTTCTGATCGAACACTAGACTGTTACACCAACCTGTTCCAGTACATC 60  
DB 1370 ATGCGACCGCTTCTGATCGAACACTAGACTGTTACACCAACCTGTTCCAGTACATC 1429  
QY 61 GCGGAGCCACCGGGCAACCCCTCGTGTGCCAAGACG---CCGACTGTGCGCG 117  
DB 1430 TCCCGGGCCACGCGCGGAGACCCGAGTCATCCGACAGACGACCCGCTGGCGGCG 1489  
QY 118 CTGCGCGTGAAGACTTTCAGACGATTCGTCGTCCCGGGGCCCGGACGCCGACCGG 177  
DB 1490 GGTCTGCTGACGCGTTCGACAAAGTGCTCTCCCGGGGCGCGGACCCCGACCGC 1549  
QY 178 GAACGGACCTTGGATCAGCCGCGGGGATCACCAGACGAGCGCTGCGCTCGCGC 237  
DB 1550 CCGGCGCACTTGCGCTGTGCGCTGCGCGGATCGCGGACGAGAGGCGCGCTGCGTCCG 1609  
QY 238 GTCTGCTTGGCCACACGAGGCAATCGCCACGCTTTCGCGGAAACGTCGCGCTCGCCCG 297  
DB 1610 GTCTGCTTGGCCACACGAGGCAATCGCCACGCTTTCGCGGAAACGTCGCGCTCGCCCG 1669  
QY 298 GAACCCATGACAGCGCGCGGTCTCCGAGTGGCGGACACCGGAGAGAGCTTTCCGGGGC 357  
DB 1670 GAGCCCGCTGACGCGCGCACTTCGCGGATGCGGACGAGGACCGGCTTTGAGGGG 1729  
QY 358 CTCCCTCGCGCTTACACGCGCGTGCCTACACTCCCTCGCGGACCGACCTCCCGGAC 417  
DB 1730 CTGCGCGACCGCTTGAAGTGTGTGGTGCATCTCCCTCGCGGTACGGAATCGCCCGC 1789  
QY 418 GAGCTGAAACCCCTCGCTGAGACGACGCGGGTGTCTATGAGGCTTGCGGACCGCGAG 477  
DB 1790 GAGCTGAGAGGACACGCGCTGTGCGAGAGCGGGTCTGATGCGCTGCGGACCGCACG 1849  
QY 478 AAGCGCGTGGGGGCTCCAGTTCACCCCGAGTCCATCGGAGAGACTTGGCCGCGGAG 537  
DB 1850 CTGCGCTGTGGGAGTGTGATTCACCCCGAGTGTGAGCACCCAGAGCGCACCGG 1909  
QY 538 ATCATGGCCAACTTCGCGACCTCGCGCCACACCGGAGCACGCG-----586  
DB 1910 CTGCTGGCCAACTTCGCGACCTTCACCGAGCCACCGGCGCAACCGCGCGGCGG 1969  
QY 587 -----GCCACGGGGCGGACTCCCGGTACGAACTCCAGTGCGC 625  
DB 1970 GCGGAGACGCGACCGCTCCCGCGCGCGCGCGCGCGGAGACGAGGCCACACCGG 2029  
QY 626 GCGTGAAGTGTGCG-----GAGCGCGAAG 654  
DB 2030 ACCCGACGCGGCTCGGATTCGCAAGTGTGCTGCCACGCGGTGGAGCCCGAGGTTC 2089  
QY 655 GTACCGCGGCGTGTGCTGCGCGCGGAGGACACAGTCTGGCTGAGACGAGCTCGTTC 714  
DB 2090 GCGTTCGACTGCGCTTTCGACCGGACACACCTTCTGCTGACGACGAGCGTCCC 2149  
QY 715 CTGGAAGCGCTCGCGCTTCTCTCTTCGCGACGACCGCGGCGCTGCGCGAGTAC 774  
DB 2150 GGGGCGAGCTGGCGCAGCTCTCCATGATGGCGACGCTTCAAGTCCCTCGCGCGGAC 2209  
QY 775 CTCACCTTACCGGTGCGCGAGCGGCGTCTTCCTTCGCGGCTCGGACGACGACGAC 834  
DB 2210 GCGAAGGCGGAGTGTGACAGCGCGAACCCTTCACGCTGAGAG---CCGAGCGCGCCGACG 2266  
QY 835 CGGAGCGGCGCGCTTCTTCAACTACCTGAGAGGAGAGCTGMAACGCGGAGCGGTCCC 894



Db	2267	ACGGTCAGAGACGCGCTTCTCTGACCTGCGCTGGAGAACGACCTGGCGGGGCTGCG---CACG	2323
Oy	895	GTGCGCCCCGGAATGCGCCCTTTCGAGTTCAACCTTCGGCTACGTCGCGTACCTTCGGCTACGAG	954
Db	2324	GAGGTCGCCGAATCTTCCCTTCGCGTTTCGCGCTCGGCTGGGTTCGGCTGCTGGGCTACGAG	2383
Oy	955	CTGAAGGGGAGACACACCGGCGGACCCCGGCGACCGGATCCCGGACCCCGACCGCGCGCTTC	1014
Db	2384	TTGAAGGGCCGAGTGGAGGAGGACCGCCCGGACCGGCTGCGCGATTCGCGAGCGCGTGTCT	2443
Oy	1015	CTCTTCGCGCGACCGCGCGCATGCGCCCTCGACCAACGAGAAAGGTGTGTCTACTCTGTGGCC	1074
Db	2444	GTCCTTCGCGGACCGGGGCCCTGCTGTGGAACCAACCGCACCCGACCACTTACTGTGTGGCG	2503
Oy	1075	CTCGACCGGCGGCGCAACGACGCGGCCCGCGGCTTGGCTTGGGAGAACGCGCGAGACG	1134
Db	2504	CTGGTCGAGGAGACGACGCGGAGGCGGAGCGGCGGCTGTGGCTTCGGGGCGGCTTCGCGCACG	2563
Oy	1135	CTCACCGGCGCTGGCGCGTTCGCGGCCCGGCGGACCGACCCCGCCAGTGTTCGGGATC	1194
Db	2564	CTGGACGCGCTGTGCGCGGGCGGGAGCGC-----CGAGCGGTGC	2599
Oy	1195	CCCGAGGCGGGGGCGCGCTTGGCGCCCGCGCGCGCGCGCGGCGGCGACGACAAAGAGCGCTAC	1254
Db	2600	CCCGAGGCGCGCGGTGTGTGACGACGCGGTTCGGTGAGCTGCGCGCACGACCGGAGCGGCTAC	2659
Oy	1255	CTCAAGCGCATCGACGAGTGCCTTCAAGAGAGATCCGCAACCGGCGAGTCTACGAGATCTGC	1314
Db	2660	CTGAAGCGTGAATCGACGTCTGCGACGAGGAGATAGCGCGCGGGAGAACCTTAGAGGTCTGCG	2719
Oy	1315	CTGACCAACATAGTGTACCGCGCGCGACCGAGGCGACGCGCTTGGCGCTTACTCTCGCGCTG	1374
Db	2720	CTGACCAACATAGGCGGAGCGGAGCACCGACCTCACCCGTGGGGCGGCTTACCGGCGCGTG	2779
Oy	1375	CGCGCGATCAGCGCCGTCCTCCGTTACGCGGCCCTGCTCGAGTTCCCGGAATCTGCGGTGCTG	1434
Db	2780	CGCGCGGTGAGCGCCCGGCCGCTTGCCTGCTTCTGGACTTTCGCTCCANTGCGGTGCTC	2839
Oy	1435	AGCGCTTCGCGCGAGCGGTTCTTCAACATTCGGGCGCGACCGCGCGGTGCACTTCAAGGCC	1494
Db	2840	AGCAGCTCTCGGAGCGGTTCTCTGGGATTCGACCGGACCGGCGGATGGAATTCGAAGCG	2899
Oy	1495	ATCAAGGGGAGCCCGCCCGCGGCGGACCGCGGAGGAGACGACGCGCTTCGCGCGGAC	1554
Db	2900	ATCAAGGGGACGCGGCGCACCGCGCGCACCCCGGAGGAGACGCGCGCTGTACGTGTC	2959
Oy	1555	CTGGCGCGGCGGGAGAAAGGACCGGGCGGAACTGATGATCTGTGACCTTGTCCGCAAC	1614
Db	2960	CTGGCGCATCTCGAAGAAAGGACCGGCGCGCAAGACCTGATGATCTGTGACCTTGTCCGCAAC	3019
Oy	1615	GACCTCAACAGCGTCTGCGGATCGGCGTCCGTCACGCGCGCCCGGCTTTCGAGGTGGAG	1674
Db	3020	GACCTGGGGCGGTGCGCGGAGGTGAGCTCGGCTCGTGTGCGGACCGGATGTTTCAAGTTCAG	3079
Oy	1675	ACCTACGCGCGCGGTGACACAGCTGATGTGACCAATCCGGGGACGGCTTGGCGCGCGCACG	1734
Db	3080	ACGTAACGGGACCGGTGACCAAGCTGGTCAGTACCGTACGAGGGCGGCTTGGGAGACGACG	3139
Oy	1735	AGCACCGCGCGCTGTGCTGACGCGCGGCTTCCCGCGGCGCTTCATGACCGGCGCGCAAG	1794
Db	3140	AGCCCGGTGGCGGCGGTTCGGGCGGCGCTTCCCGCGGCGGTGAGTAGCGGCGGCGCGAAG	3199
Oy	1795	AAGGCGACATGAGGATCATGACCGGCTGGAGGAAAGGCGCCCGGGGGGTTCATCCGGGG	1854
Db	3200	ATCCGACATATGACGATATGACCGGCTGGAAAGCGGGCGCGCGGTGTCTACTCTGGGCG	3259
Oy	1855	GCGCTCGGATGATTCGCGCTCAGCGCGCGCGCGGACCTCAGCATCTGTATCTCCGACCATC	1914
Db	3260	GCGATTCGGCTACTTCTCCCTCACCGGGCGGGTAAAGACGTGCAATGTGATCTCGCACGGGT	3319
Oy	1915	GTCGTCGCGGACGCGCGAGGAGTTGGGCGTTCGGCGGGGCGGATGTGTCTCTCTCGGAC	1974

Db	3320	GTGCTCAGCGCGGCGAGACTGCGCTACGGGCTGCGGCGCGCCGCTCANTCGGCTCTCCGAC	3379					
Qy	1975	CAGAGAGAGATTACCGAGACCGCTGTAAAGCCCGCGCCATGTCACCGCCCTTGAC	2034					
Db	3380	CCGGCCGACGAGTTTCGAGAGACGGCGCGCTCAAGCGCCCGCGCTGCTGCTCTTCGAC	3439					
Qy	2035	GGCAGCGCGCTGCGCGGCGCGCCG	2057					
Db	3440	ACCGCCTTCCCGGCGCGTGACGC	3462					
RESULT 9								
SGR300302	39314 bp	DNA	linear					
BCT 20-MAR-2002								
ACCESSION	Streptomyces griseus partial ORF1, canR gene, canC gene, canF gene, canT gene, canR gene and canRB gene.							
LOCUS	AJ300302							
DEFINITION	AJ300302.1 GI:12231153							
KEYWORDS	ABC-transporter; aminotransferase; can RB gene; canR gene; canC gene; canF gene; canR gene; canT gene; Cho-like protein; cytochrome P-450; ferredoxine; PABA synthase; thioesterase.							
SOURCE								
ORGANISM								
Streptomyces griseus								
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.								
REFERENCE								
1								
Campelo, A.B.								
Thesis (2000) Department of Microbiology, Universidad de Leon,								
Leon, Spain								
REFERENCE								
2								
Campelo, A.B. and Gil, J.A.								
The candidicin gene cluster from Streptomyces griseus IMRU 3570								
Microbiology (Reading, Engl.) 148 (Pt 1), 51-59 (2002)								
JOURNAL								
MEDLINE								
21642576								
PUBMED								
11782498								
REFERENCE								
3 (bases 1 to 39314)								
Gil, J.A.								
Direct Submission								
Submitted (14-NOV-2000) Gil J.A.; Microbiologia, Universidad de								
Leon, Campus de Vegazana s/n, 24071, SPAIN								
FEATURES								
source								
1..39314								
/organism="Streptomyces griseus"								
/mol_type="genomic DNA"								
/strain="IMRU 3570"								
/db_xref="taxon:1911"								
complement (<1..1785)								
/note="ORF1"								
/codon_start=1								
/transl_table=1								
/product="Cho-like protein"								
/protein_id="CAC22112.1"								
/db_xref="GI:12231154"								
/db_xref="UniProt/TrEMBL:Q9WC6"								
/translation="MLERTTEIATVDALGAATEGRSSILVTGPIGSRALLORL								
SCIARDVDVRLRAHAAMEQDFAGVQLFETLLGDSPEADPRMDAHAFARH								
VLADAPPGADQAIATAEVATLGHLSILANVSAQDSRLILVDDLGSDVPSRLITY								
LARKHGAIVVVCALRDGPRSHHTLVREIAPAAQTLPASTSLATHEIVREHNG								
EADDERVQACHASVGNPFLISILVGTGFLGRPLAEATARRRRPQLERLAS								
ILITQPPVADILAAIATLGEQSDAPFLALGLDSTIGVAGLRAAGALGALAPDEP								
RITRSPVDAESTLTVMVQERHMDAAALVYAGSPABVQAQLVAVTRPQWADV								
VLSAADTLARGAPDTAAGYLRCALLDSFAACVGRLLVLEGTBERGDPPLACERH								
IAAMNTLPEPRDAVAAIRISPTALGPALTVDLLROAEDLGADGLGTDRDLA								
LRLERLRHGHGHEDPHEILASARHLKEIGPEPVDVGGRELLITALGAATLGHPLA								
SEVRLRAGRYLDEPATSRIHSTPLAT"								
9525..10583								
/gene="canA"								
/gene="canA"								
9525..10583								
/gene="canA"								
/codon_start=1								
/transl_table=1								
/product="aminotransferase"								
/protein_id="CAC22113.1"								



/db\_xref="GI:12231155"  
/db\_xref="GOA:O9EMC5"  
/translation="MAFTHPVSRPALDGRRELYVSDAVSGGWTSSOGPVYRPREEAP  
ENMGVAGVACSSGTALTLALRALNIGDEDEVIVPEFTMVASAMAVTYTGATPVVAD  
CGDDLNIDVTRIEBKITPTTRAVMPVHVTRCKDMQVMDLALQYNLRVEDSEAHG  
VRVGDIAFCSLPAANKITVAGEGVCLTDQRLAEQOLAHARAFTRDSFLMKLIAR  
NYMTAMQACVALAOTERLDEILATREIELARYDAGLKDLPGITLMPARDVLMYDLR  
AERREIELRAHLDERGIETRLFFKPMRSQPYLDPPVPTLNAHRSSEDLGYLPTHGTLT  
AADOEYITGARDVPFYRAG"  
10678 . 11859  
/gene="canC"  
/gene="canC"  
10678 . 11859  
/gene="canC"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P-450"  
/protein\_id="CAC22114.1"  
/db\_xref="GI:12231156"  
/db\_xref="GOA:O9EMC4"  
/db\_xref="UniProt/TREMBL:O9EMC4"  
/translation="MTTSPGTVVDPPTPTPREPLPSOYAEHRKONGLVOTHLPNCR  
PIWLVRHEBYRAVLTHPRISANPDNPGPNVGETNGVKOEQIPGMFGLDSEPHDR  
ERKVLPEFTVRVRELRLPAIERTVBERIDAMLAGNTADLVNDFALPVPSLVIISLL  
GVSADRPFPESRTKLVAIRISIDBERAATQOLRYINRLVIKKKRGEDLISRL  
LSTGKLSDELSGVLLLIAGHETTANNIGLVVTLISHEWIGDDRVLVEELLRLHS  
VADVMALRVAVDVEIAGOTIRKGEIVPLASANDTEAFGCPHAFNPRTERRHYA  
FGYGVHQCIGQNLVVRVEMELAYRKLPERIDELALVPEQOLAKYDGIIFGLHELEVR  
W"  
11906 . 12100  
/gene="canF"  
/gene="canF"  
11906 . 12100  
/gene="canF"  
/codon\_start=1  
/transl\_table=11  
/product="ferredoxine"  
/protein\_id="CAC22115.1"  
/db\_xref="GI:12231157"  
/db\_xref="GOA:O9EMC3"  
/db\_xref="UniProt/TREMBL:O9EMC3"  
/translation="MRYTVDSBQCVAGAGCVLNAPEVFPDDDDGVVLLRAEPDERDH  
EAVRIAGDLCPSSASVVLQBD"  
12223 . 12993  
/gene="cant"  
/gene="cant"  
12223 . 12993  
/gene="cant"  
/codon\_start=1  
/transl\_table=11  
/product="thioesterase"  
/protein\_id="CAC22116.1"  
/db\_xref="GI:12231158"  
/db\_xref="GOA:O9EMC2"  
/db\_xref="UniProt/TREMBL:O9EMC2"  
/translation="MTTADDTAARMLRHRHPAADAIVLVCFPHAGSASFYHVPYSAR  
FAGAEVYSIQYGRQDRKEPCVPLGTADLITBQLLPLDERPTVFGHSGSALIA  
FETAMLEQKAGPRTVIASGRGSPSTTAEARVHTDDGIVEMKRLVINGTAGVGLAD  
BEILRALPALRGDYRAIETTCPPDRILACGLTVLTGEDDPLTTVEEAERMDHTTG  
PFLRLVFTGGHFFELTQLDAVNTIEIALHPDAARA"  
13190 . 15361  
/gene="pabAB"  
/gene="pabAB"  
13190 . 15361  
/gene="pabAB"  
/codon\_start=1  
/transl\_table=11  
/product="PABA synthase"  
/protein\_id="CAC22117.1"  
/db\_xref="GI:12231159"  
/db\_xref="GOA:O9EMC1"  
/db\_xref="UniProt/TREMBL:O9EMC1"  
/translation="MRTLLVDNYDSFTYNIPLYLSRANGREPEVINDDPAMRPGLLD  
AFPNVYLSREPGRPHRPAIDGICARIAEBRCLPVLGVCIHQOMALAHARVGRAPED  
RHRGRTAVRHDTGTFEGLPQPLEVVRVYHSLAVTELPELELAWSDEDLVALHART  
LPLMGVQFHPESIGTODGHRLLANFRDLTERHGRTBHGRRAGHTLPPAPARETVAT

TCGPRRLVIAKSLPTRMDEAVAFDLSLFRITGDHPWMLDSRREGELGQLSNMGDASGP  
LARIKAADAGACTVTRADGASSTVESALVWLDLRTVEPELPPAPALGMWGC  
LYEILKAEDDGAARSRPDDAVLVPAADALVLDHRTTLYTLVVEDDAEABARAWL  
LAASATDLVAVAREBPCEPCREAPVCTTGPVELRHDDGYLKLIDVCOQELIACTEYVC  
LTNMBADTLVPMAYRALRVSPAPFALIDFGMAVLSSEPFLTIDHGRMES  
KPIKTRPRGATPPQEDALVRLALACEKRAENLMI VDLVRHDLRCAGVGVVADPV  
PQVETVATVHQVLSVTYALREDSPVAVRAPFGSGMTGAPKRTMOIIDRLGEP  
RQVYGAIGYFELTGAVDLSIVIRVYSGGRRLRYGVGAVIALSDPADEFRETVAKA  
AFLRLDLDPGPREAPGKDLDBGDDGDGADAPDVLVPG"  
20829 . 21836  
/gene="canB"  
/gene="canB"  
20829 . 21836  
/gene="canB"  
/codon\_start=1  
/transl\_table=11  
/product="ABC-transporter"  
/protein\_id="CAC22118.1"  
/db\_xref="GI:12231160"  
/db\_xref="GOA:O9EMC0"  
/db\_xref="UniProt/TREMBL:O9EMC0"  
/translation="MNGIHLPPAPLPAPVCSAAGSGDLMTEAGVSKAYGTVAHLDV  
SLAVRGTYGLLGHNGAKTTLVILITLALPPTGRRARVAVDYADKPEVRRRIGL  
TQCFASVDAQLSGYDNVLILALDAGRRAPVRLRELELFRLEVDARRASSYSG  
LRRRLDLAVSLVAGBEVLFLDEPTGLDPSRINIMEIVELVBOGTVLLTVOYLEB  
ADRLDLAVLSAGVVVAGTAPELKATVGRVTTLTLETGDSVAARTALRAGAFAP  
VGEARVTVVPIDATREIADIIRSLDQAGVEASELNFGEPTLDVYTLAEGAAKTR  
A"  
21970 . 22776  
/gene="canB"  
/gene="canB"  
21970 . 22776  
/gene="canB"  
/codon\_start=1  
/transl\_table=11  
/product="ABC-transporter"  
/protein\_id="CAC22119.1"  
/db\_xref="GI:12231161"  
/db\_xref="GOA:O9EMB9"  
/db\_xref="UniProt/TREMBL:O9EMB9"  
/translation="MPRAVAPAPFGSSILFTQIRVLGTRSLRAMVTDPEIVLFGIOP  
VILVLVLDVPEKMGMPHFPDGVYLDIVLVAVLVDNAAGSMGSGVGLVDDKNGI  
VARLSLPPVHGPALLAARSLVGLVRSAAQVATMLAMTVLGSISNGAABIALSAGL  
TLFISFLGMAPIAAGAMLRAPLQNLALIVIFPLMPSASVYVADLPVLAGAVAS  
VNPLTVADATRALADVGLDHAIPAVVICAVIIVGMIAVARGFRPL"  
ORIGIN  
Query Match 37.5%; Score 773; DB 1; Length 39314;  
Best Local Similarity 63.3%; Pred. No. 8.5e-55;  
Matches 1337; Conservative 0; Mismatches 680; Indels 96; Gaps 6;  
QY 1 ATGGCGAGCGCTTCTGATCGACAACGACTGTTCAACCACACCTGTCAGTACAC 60  
DB 13190 ATGCGGACCTTCTGTCGACAACTAGACTGTTCACTTCAACCTCTTCACTACCTC 13249  
QY 61 GGCAGAGCCACCGGGGACACCCCGTCTGTCGCCAACGACG---CCGACTGATCGCGG 117  
DB 13250 TCCCGGGCCAAAGCGCGGGAACCCGAGTCATCCGGAAGACGACCCGCGTGGGGCGCG 13309  
QY 118 CTGCGCGGTGAGAGACTTGCAGCGGATCTGTCGTGTCGCCGAGCCCGGACGCCGACCGG 177  
DB 13310 GGTCTGCTCGACGCGCTTGCACACGATGAGTCTCTCCCGGGGCGCGGACCCCGACCCG 13369  
QY 178 GAAACGGGACTTGGGATCAGCGCGGGGAGATCAGCGAGCGGCGTGGCTGCTCCGCGG 237  
DB 13370 CCGGCGGACTTGGGCTGTGCGCCCGGATCGCGAGAGAGGCGGCTGGCGGTGCTCGGCG 13429  
QY 238 GTCTGCTCGGCGCACCGAGGCAATCGCCAGCTTCTGCGGGAACCGTCGCGCTCGCCCGCG 297  
DB 13430 GTCTGCTCGGCGCACCGAGGCAATGCGCCCGCCACCGGCGCCGGGTGGGCGCGGCGCCC 13489  
QY 298 GAACCATGCAAGCGCGGCTCTTCCGAGGTGCGGCAACCGGCGAGAGAGCTCTTCCGGGGCG 357  
DB 13490 GAGCCCGGTCACGCGCGGCACTTGGCGGATCGGCGGACGACGCGACCGGCGGCTCTTCAAGGGG 13549







Shanghai Jiaotong University, 1954 Huashan Road, Shanghai 200030,  
China

## FEATURES

source

location/Qualifiers

1. 138203

/organism="Streptomyces sp. FR-008"

/mol\_type="genomic DNA"

/db\_xref="taxon:206662"

complement (574. .1950)

/gene="fscO"

complement (574. .1950)

/gene="fscO"

/note="putative FAD-dependent monooxygenase"

/codon\_start=1

/transl\_table=11

/product="FscO"

/protein\_id="AA082549.1"

/db\_xref="GI:34766436"

/translation="WPSTCPAPTSIGRGCPHRRKRTSGRRRIAVPTRDCGVKRGH  
QAKKRVSMRGSVAVVGSIAGCAALASRGARITVFERADARLRDVGIGV  
HDAVYELRDYAGADDEMPWAPLNRVYVSVDGAEGRATGTPFFRAYNMGSLMS  
ELRRVPEATYRAGAKVEAEODADGVTRLADGPEREDLVIGADYRSVVRAMV  
PGTAAVAGYIGMRGTSPDVSGLPDGLDAHNTTFPGHGMAYRIDPGSGHRLNWL  
YTAARRIDGHPDLRTPTSLPRGLNAELTEHRLALVAEFPFPMAKLSTAEITF  
IQPTLDVPHYATDRNVLIGDASVARPHLGSVYALQDATALLEAMVAGSWEKI  
LEGTHARGPVTAMVGLARMGSAQVEETPDWSAMQAEFDAMWQENNGSDRSGF  
GSHLSKR"

complement (2264. .3037)

/gene="pabc"

complement (2264. .3037)

/gene="pabc"

/note="putative 4-amino-4-deoxychorismate lyase"

/codon\_start=1

/transl\_table=11

/product="Pabc"

/protein\_id="AA082550.1"

/db\_xref="GI:34766437"

/translation="MIELDEPPAGPEALASIALTNGHFTLLVENGVRGLDLHER  
LIRDCRPLFPAALDPRVRKLARAAPTDGRATVRVFPALNLGINADAPGILV  
TSRAPDKPGPLRVSVRHRLPEVKSGLCTPLLRQAOAGYDVLFTGPDGD  
ILEGTTNVLVRGEVYVWPGEVTLACTROLIRGRHSRHRTPRGPRLSGGLRD  
QRVGVPRVTGIDDRFPPAAHSTRLAEITYQALPGSPV"

complement (3150. .3818)

/gene="fscRI"

/gene="fscRI"

/note="putative transcriptional activator"

/codon\_start=1

/transl\_table=11

/product="FscRI"

/protein\_id="AA082551.1"

/db\_xref="GI:34766438"

/translation="MDPAPADAPGSPERAALVADGFRAGAVVACLPSTLTOVNO  
EERRRPGSSSELGSRFCDLVHRSVQOPLMHOPARMLDKRRHFATYVAUOERT  
STPLNALAVRGKTPDPAALLVYMNABEDADWAPRKLLSTIDRILLEGIA  
GVSTPLASRLYLSRQGEVYHVTGLRKLVPKRAALVSRASVGLKGTWPKYVE  
DFIK"

complement (4377. .7205)

/gene="fscRII"

/gene="fscRII"

/note="putative transcriptional regulator"

/codon\_start=1

/transl\_table=11

/product="FscRII"

/protein\_id="AA082552.1"

/db\_xref="GI:34766439"

/translation="MLERTTEALVDAALGAATEGRNSLVLLTGPLIGORSALQRL  
SLADDDVRLVAHAAPMEODFAFVSVQLFETLLGSDPEDARDMDHASFAH  
VLADAPPGADOLIAATEAVILGILSLANVSADSLITLVLDLQMSDPSLRMLTY  
LAKRIHGLRAVVCALDGDGPRSHTLARIRAGCOTTLAPASISLRTAHEVREHG  
EAGDDEFVQACHEASVGNPLFLUSILVGTGFLGRPLAEHETARRLRPSQLERLAS  
ILRTQPAVRDLAAIAILGEGSDAPLALAGDSIGVAGALRALGALGALAPPEP  
RFHRSVDAAEETLTWORERMHDEAALLVYAGSPAEQVAQALMAVTVTRRPPWAD

gene

CDS

/gene="fscRII"

complement (7210. .10254)

/gene="fscRII"

/note="putative transcriptional regulator"

/codon\_start=1

/transl\_table=11

/product="FscRII"

/protein\_id="AA082553.1"

/db\_xref="GI:34766440"

/translation="MYDLHMPPIELPLERDELEALSAVIGELSGRPVAVTVGEP  
GICNDLILMMAAYATDAGTRVLSAATPAHEHRYGVVAQILAEENRLAARLELT  
EORGCLPGNLGNVARDPPTLLVRDTOMUDPASIRMLERARLPRAVALUTST  
GTALTRPENSVPPLTGLATTLELPLUTSGTATVAVRATGADGDEFTALAEAT  
RGIPAVHVDLDRFARAGSPRADRIDELARLTAEVGDHATRALDLDPAVALAR  
ALAVCGDLDPFLVCTIAGPHSVSRSLAALASGLTTLRDGHRPVODAVVRAVLE  
EMPADRAELTVARAAGLAQVADOGIDLLILLRPVDPDAVDTLDFGFSALRG  
RDDIAVAVLARALDEPLAEDRARIETFOIASVEMQTPPAARERLGGILIRAPRG  
GIRARATDCLIGGTTRAAHRLAGATIDAPAAPPRRPTGAPAVPREAPASPT  
HLVPHRYPVATGPGATGPDGRSAAPADITGTRADHAPSPGEEAGRDYAGPDEP  
VAAERHELTVLCRVASFLRHDDADLDVLPVPLPAVNHAPVAGVAMHGVAGVIRE  
EVARLERSAFVPAATGRPLLVPRLLGCPALLTDDGAEHNLGLLDSHERRAI  
SVAHILTVRAELHIRHGRSDATTAQALPFRGVLATPDRSPHQRHEFRACGWWLHG  
HTEARAEATARPLPLAHSATTQOLPFRGVLATPDRSPHQRHEFRACGWWLHG  
CANPAVQPMRSLAABAAHTLGDTTEBARLVHEVRILARKWCAAPELGRQSLAVYTE  
ENRVENTLRAAVTLLSASARTYATRAVLELAABEENRRRTAVALGPATVLPVAP  
PGPHVPSVAVRPPGPRPAVLGTRKARPGHDAVWMGLTVERBTAALAAQGLN  
EIALEIAVTTRAVELLSGVYRKLIRRGHEURLVQAEAGS"

complement (10298. .13315)

/gene="fscRIV"

/gene="fscRIV"

/note="putative transcriptional regulator"

/codon\_start=1

/transl\_table=11

/product="FscRIV"

/protein\_id="AA082554.1"

/db\_xref="GI:34766441"

/translation="MWLRWAGDRSRRGAGRSAGSPSLASDARESEVPHRSRAP  
PAPPLVGRAHQDLVLARHADARAGRPRLVLDGPAIGKTAALLAALDEGPLAGMT  
TLYGSCRAVDVATGYSGRVALFGGGLTVGRKGTSPGLVGGRRALPALAOPGELDA  
DGSPTPSVIOGLYMLAANMADGPIVLIVDVLGTDGDESLRATLADGLPLFV  
LAHRTTGCTATPDAALDIYAHNLPASLGCPIDTAEVAILTHSPFEDDPLSPFGR  
LSVNTGSPLEIVRLRLDLRSAGLGPDDTALAAQGVVAASVAGVLEHDPDWRO  
VARAVSLGDEBDRITYLAAGRSVVHVEBAVILIDPAGLHGHVELSHDLAQAARV  
AVAGSGVAVLRRRAALISDGRSPEELITQOLLPGTDPDVAVSLRDAQAQLEORG  
ACEAARVLEVRBAEPKDPVLSRGLKATLDPARSVTLHBAHSLTTDVARAAT  
AVOGLTCLANVOOSPDGARALTEALDADLDELGPREEPRDRELRTLSALLIVSDE  
KATLPDILRTGTLTPOEDTPAQROQLMNLVLSABEGCAEBTVGRARRLRAPGV  
PLGVMSLALPTSLALSDENEAEEVLETVLSSGSGTAAMVTVLALSTRSLFRLNG  
AVPDMADAQTALEITIGERMGDTLTPHTAASVYTEGDEGRALDALDAKRPILD  
RFVWEYHWYLMARGALADGDLGALGVSGSGAMAGLTNPVLAWMLETACLL  
GEAGREGGEAGRAAGTTRLAERWGRBALVGAVALRGAARGTARTGALREAVPLAD  
SPARSGHARGILLRLAIVEDAVNREGRHLEAENGLARCCCVALARARABELIAG  
GRREYVASPLMDLTGTERTVAGLVASGNEBVBESLFTVTVLEHLITSYRKLG  
ARGDITFALRAGATARPAAERRPCHAKRNP"

13522. .14898

/gene="fscMI"

/gene="fscMI"

/note="glycosyltransferase responsible for mycosamine  
attachment"

/codon\_start=1

/transl\_table=11

/product="FscMI"

/protein\_id="AA082555.1"

/db\_xref="GI:34766442"

/translation="MWLRWAGDRSRRGAGRSAGSPSLASDARESEVPHRSRAP  
PAPPLVGRAHQDLVLARHADARAGRPRLVLDGPAIGKTAALLAALDEGPLAGMT  
TLYGSCRAVDVATGYSGRVALFGGGLTVGRKGTSPGLVGGRRALPALAOPGELDA  
DGSPTPSVIOGLYMLAANMADGPIVLIVDVLGTDGDESLRATLADGLPLFV  
LAHRTTGCTATPDAALDIYAHNLPASLGCPIDTAEVAILTHSPFEDDPLSPFGR  
LSVNTGSPLEIVRLRLDLRSAGLGPDDTALAAQGVVAASVAGVLEHDPDWRO  
VARAVSLGDEBDRITYLAAGRSVVHVEBAVILIDPAGLHGHVELSHDLAQAARV  
AVAGSGVAVLRRRAALISDGRSPEELITQOLLPGTDPDVAVSLRDAQAQLEORG  
ACEAARVLEVRBAEPKDPVLSRGLKATLDPARSVTLHBAHSLTTDVARAAT  
AVOGLTCLANVOOSPDGARALTEALDADLDELGPREEPRDRELRTLSALLIVSDE  
KATLPDILRTGTLTPOEDTPAQROQLMNLVLSABEGCAEBTVGRARRLRAPGV  
PLGVMSLALPTSLALSDENEAEEVLETVLSSGSGTAAMVTVLALSTRSLFRLNG  
AVPDMADAQTALEITIGERMGDTLTPHTAASVYTEGDEGRALDALDAKRPILD  
RFVWEYHWYLMARGALADGDLGALGVSGSGAMAGLTNPVLAWMLETACLL  
GEAGREGGEAGRAAGTTRLAERWGRBALVGAVALRGAARGTARTGALREAVPLAD  
SPARSGHARGILLRLAIVEDAVNREGRHLEAENGLARCCCVALARARABELIAG  
GRREYVASPLMDLTGTERTVAGLVASGNEBVBESLFTVTVLEHLITSYRKLG  
ARGDITFALRAGATARPAAERRPCHAKRNP"

13522. .14898

/gene="fscMI"

/gene="fscMI"

/note="glycosyltransferase responsible for mycosamine  
attachment"

/codon\_start=1

/transl\_table=11

/product="FscMI"

/protein\_id="AA082556.1"

/db\_xref="GI:34766443"

/translation="MWLRWAGDRSRRGAGRSAGSPSLASDARESEVPHRSRAP  
PAPPLVGRAHQDLVLARHADARAGRPRLVLDGPAIGKTAALLAALDEGPLAGMT  
TLYGSCRAVDVATGYSGRVALFGGGLTVGRKGTSPGLVGGRRALPALAOPGELDA  
DGSPTPSVIOGLYMLAANMADGPIVLIVDVLGTDGDESLRATLADGLPLFV  
LAHRTTGCTATPDAALDIYAHNLPASLGCPIDTAEVAILTHSPFEDDPLSPFGR  
LSVNTGSPLEIVRLRLDLRSAGLGPDDTALAAQGVVAASVAGVLEHDPDWRO  
VARAVSLGDEBDRITYLAAGRSVVHVEBAVILIDPAGLHGHVELSHDLAQAARV  
AVAGSGVAVLRRRAALISDGRSPEELITQOLLPGTDPDVAVSLRDAQAQLEORG  
ACEAARVLEVRBAEPKDPVLSRGLKATLDPARSVTLHBAHSLTTDVARAAT  
AVOGLTCLANVOOSPDGARALTEALDADLDELGPREEPRDRELRTLSALLIVSDE  
KATLPDILRTGTLTPOEDTPAQROQLMNLVLSABEGCAEBTVGRARRLRAPGV  
PLGVMSLALPTSLALSDENEAEEVLETVLSSGSGTAAMVTVLALSTRSLFRLNG  
AVPDMADAQTALEITIGERMGDTLTPHTAASVYTEGDEGRALDALDAKRPILD  
RFVWEYHWYLMARGALADGDLGALGVSGSGAMAGLTNPVLAWMLETACLL  
GEAGREGGEAGRAAGTTRLAERWGRBALVGAVALRGAARGTARTGALREAVPLAD  
SPARSGHARGILLRLAIVEDAVNREGRHLEAENGLARCCCVALARARABELIAG  
GRREYVASPLMDLTGTERTVAGLVASGNEBVBESLFTVTVLEHLITSYRKLG  
ARGDITFALRAGATARPAAERRPCHAKRNP"

13522. .14898

/gene="fscMI"

/gene="fscMI"

/note="glycosyltransferase responsible for mycosamine  
attachment"

/codon\_start=1

/transl\_table=11

/product="FscMI"

/protein\_id="AA082557.1"

/db\_xref="GI:34766444"

/translation="MWLRWAGDRSRRGAGRSAGSPSLASDARESEVPHRSRAP  
PAPPLVGRAHQDLVLARHADARAGRPRLVLDGPAIGKTAALLAALDEGPLAGMT  
TLYGSCRAVDVATGYSGRVALFGGGLTVGRKGTSPGLVGGRRALPALAOPGELDA  
DGSPTPSVIOGLYMLAANMADGPIVLIVDVLGTDGDESLRATLADGLPLFV  
LAHRTTGCTATPDAALDIYAHNLPASLGCPIDTAEVAILTHSPFEDDPLSPFGR  
LSVNTGSPLEIVRLRLDLRSAGLGPDDTALAAQGVVAASVAGVLEHDPDWRO  
VARAVSLGDEBDRITYLAAGRSVVHVEBAVILIDPAGLHGHVELSHDLAQAARV  
AVAGSGVAVLRRRAALISDGRSPEELITQOLLPGTDPDVAVSLRDAQAQLEORG  
ACEAARVLEVRBAEPKDPVLSRGLKATLDPARSVTLHBAHSLTTDVARAAT  
AVOGLTCLANVOOSPDGARALTEALDADLDELGPREEPRDRELRTLSALLIVSDE  
KATLPDILRTGTLTPOEDTPAQROQLMNLVLSABEGCAEBTVGRARRLRAPGV  
PLGVMSLALPTSLALSDENEAEEVLETVLSSGSGTAAMVTVLALSTRSLFRLNG  
AVPDMADAQTALEITIGERMGDTLTPHTAASVYTEGDEGRALDALDAKRPILD  
RFVWEYHWYLMARGALADGDLGALGVSGSGAMAGLTNPVLAWMLETACLL  
GEAGREGGEAGRAAGTTRLAERWGRBALVGAVALRGAARGTARTGALREAVPLAD  
SPARSGHARGILLRLAIVEDAVNREGRHLEAENGLARCCCVALARARABELIAG  
GRREYVASPLMDLTGTERTVAGLVASGNEBVBESLFTVTVLEHLITSYRKLG  
ARGDITFALRAGATARPAAERRPCHAKRNP"

13522. .14898

/gene="fscMI"

/gene="fscMI"

/note="glycosyltransferase responsible for mycosamine  
attachment"

/codon\_start=1

/transl\_table=11

/product="FscMI"

/protein\_id="AA082558.1"

/db\_xref="GI:34766445"

/translation="MWLRWAGDRSRRGAGRSAGSPSLASDARESEVPHRSRAP  
PAPPLVGRAHQDLVLARHADARAGRPRLVLDGPAIGKTAALLAALDEGPLAGMT  
TLYGSCRAVDVATGYSGRVALFGGGLTVGRKGTSPGLVGGRRALPALAOPGELDA  
DGSPTPSVIOGLYMLAANMADGPIVLIVDVLGTDGDESLRATLADGLPLFV  
LAHRTTGCTATPDAALDIYAHNLPASLGCPIDTAEVAILTHSPFEDDPLSPFGR  
LSVNTGSPLEIVRLRLDLRSAGLGPDDTALAAQGVVAASVAGVLEHDPDWRO  
VARAVSLGDEBDRITYLAAGRSVVHVEBAVILIDPAGLHGHVELSHDLAQAARV  
AVAGSGVAVLRRRAALISDGRSPEELITQOLLPGTDPDVAVSLRDAQAQLEORG  
ACEAARVLEVRBAEPKDPVLSRGLKATLDPARSVTLHBAHSLTTDVARAAT  
AVOGLTCLANVOOSPDGARALTEALDADLDELGPREEPRDRELRTLSALLIVSDE  
KATLPDILRTGTLTPOEDTPAQROQLMNLVLSABEGCAEBTVGRARRLRAPGV  
PLGVMSLALPTSLALSDENEAEEVLETVLSSGSGTAAMVTVLALSTRSLFRLNG  
AVPDMADAQTALEITIGERMGDTLTPHTAASVYTEGDEGRALDALDAKRPILD  
RFVWEYHWYLMARGALADGDLGALGVSGSGAMAGLTNPVLAWMLETACLL  
GEAGREGGEAGRAAGTTRLAERWGRBALVGAVALRGAARGTARTGALREAVPLAD  
SPARSGHARGILLRLAIVEDAVNREGRHLEAENGLARCCCVALARARABELIAG  
GRREYVASPLMDLTGTERTVAGLVASGNEBVBESLFTVTVLEHLITSYRKLG  
ARGDITFALRAGATARPAAERRPCHAKRNP"

13522. .14898

/gene="fscMI"

/gene="fscMI"



```

/translate=11
/product="Fscm1"
/protein_id="AA082555.1"
/db_xref="GI:34766442"
/translation="KDSAAKPIILFVSLPSSGLINPLVLVAGELARRGVDPDLNFAITDEH
RKEEVVALDVSFVSFASLGEVVELSAVTWDEVEVREYQSSRFPAHVAVQSYRP
ALQARKYRELQVDEVRPALMNVDCVAGFVDLARGIPLYLNPVFAVSNVLTSHN
PFGASYTPKSPVPVNSGLPARNMSROKLANITLPKRTLGMLFHPDMAALLBEDAALTR
ELGIAPRNMTVRDEAAAVCSVAELDVDPDI PDRVSLVGATLPLPEAPDDDEVTR
WLDQASVVMYMGRTITRLTRREVALVEYARMSTHQPJKLPEQOHLPEASGL
PDLRVESVPSQDLVLAHPNVSVFSSHGGNAVHGKVPKQVVRPLWVDFDQAV
RGRDFGISLTLDKPHTVDPDVVDKLTVRTSPAFRTAEARLQRAAGRAAADL
VTGLLPAA"
14953..16011
/gene="fscm1"
14953..16011
/gene="fscm1"
/feature="putative aminotransferase responsible for
mycosamine biosynthesis"
/codon_start=1
/translate=11
/product="Fscm1"
/protein_id="AA082556.1"
/db_xref="GI:34766443"
/translation="MAFTHPVSRPALDGRELEVDVAVSGNISOGPYVRRPEEAF
ENMGVHGVASGSGTALTLALBALNIGDEVIYVEFTMVASAMVTTGTGATPVVD
CDDDLNIDYTRJBEKTPRTTRAMPHYVYGRCDMDVMDLALQYTLRVBDSAEHG
VAPVDIACFSLPANKITTAGBGVCLTDPPRLABQLAHRAFAFRDHSFLHKILAY
NRRMTMQAVALAOTERLDELATRRLEIARYDAGLKDLPGITLMPADVLMYDLR
ASREBELRAHLARGLIETRLFFKEMSRQGYLDPVPTLNARHFSBDGYLPHHTLT
AADQEYITGAVRDFYRAG"

```

Query Match 37.4%; Score 769.8; DB 1; Length 138203;

Best Local Similarity 63.2%; Pred. No. 9.2e-55;

Matches 1335; Conservative 0; Mismatches 682; Indels 96; Gaps 6;

```

QY 1 ATGCGACGCTTGTGATCGACACTGACACTCGTTCAACCCAGCAACCTGTTCCAGTACATC 60
DB 18610 ATGCGACCTTGTGATCGACACTGACACTCGTTCAACCTGTTCAACCTGTTCAACCTC 18669
QY 61 GGGGAGGCGACCGGGGACACCCCGCTGTCGTGCGGACGCG---CCGACTGTGCGG 117
DB 18670 TCCCGGGCCAAACGGCGGGGAAACCGAGTCATCCGCAACGACCGCGGCTGCGGCGCC 18729
QY 118 CTGCGCGTGAAGACTTTCAGACGAGATCTGCTGTCTCCCGGGCGCCGCGACGCCGACCG 177
DB 18730 GGTCTGCTGACCGCTTTCAGCAACGTGTGCTCTCCCGGGCGCGGCTACCCCAACCGC 18789
QY 178 GAAACGGGACTTGGGAATCAGCGCGCGGGGAGATCAGCGACAGCGGCGCTGCGCTCGCG 237
DB 18790 CCGGCGGACTTGGCGCTGTCGCGCGCGAGATCGCGAGAGGGCGCGCTGCGGCTCGCG 18849
QY 238 GTCTGCTCGGCGACCGAGGAGCATCGCGACTCTTTCGCGGAAACGTTGCGCTCGCGCG 297
DB 18850 GTCTGCTCGGCGACCGAGGAGCATCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 18909
QY 298 GAAACCATGACCGCGCGGCTTTCAGAGTCCGCGACACCGCGGAGAGACGTCTTTCGCGG 357
DB 18910 GAGCGCGCTGACCGCGCGGACCTTCGCGGATACGCGACGACCGCGGCTCTTTCAGAG 18969
QY 358 CTCCCTCGCGCTTTCACCGCGCGGCTTTCACCTCTCGCGCGCGCGCGCGCGCGCGCGCG 417
DB 18970 CTGCGCGAGCGCTGAGAGGAGGAGTACCTCTCGCGGATACGCGAGCTGCGCGCGCG 19029
QY 418 GAGCTGAAACCCCTCGCTTTCAGACGACGAGGAGTGTATGAGGCTGCGCGCGCGCGAG 477
DB 19030 GAGCTGAGAGCGACCGCGCTGATCGAGAGACGGGCTCTGATGCGCTGCGCGCGCGAG 19089
QY 478 AAGCGCTGTGGGCGCTTTCAGTTCACCGGAGTCCATCGGCGAGCGACTTTCGCGCGGAG 537
DB 19090 CTGCGCGCTGTGGGCGGATTCACCGCGAGTTCATCGGCGACCGAGCGCGCGCGCG 19149
QY 538 ATCATGGCAACTTTCGCGGAGCTCGCGCTGCGCGACCGAGGCGAGCGC-----GCGAC 591

```

```

DB 19150 CTGCTGGCAACTTTCGCGGAGCTTCACGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 19209
QY 592 GGGGCGGAGCTTCCCGGATCGAACTTCACGATGCGCGCGGCTGAGCTGTGCGCGAGCGCGAA 651
DB 19210 GCGGGGCAAGGACAGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19269
QY 652 GAGGTACGCGCGGCGGCTGCGCGCGCGCGCGAG-----GCGCGCGCGCGCGCGCGCG 681
DB 19270 AGCGCAGCGCGGCTCGGGATCATTCGCGAGCTGCTGCCACGCGCTGCGAGCGCGAGGTC 19329
QY 682 -----GCGACCAACGCTTTCGCTGAGCAGACGACTCGCTC 714
DB 19330 GCCTTGACCTGCTGTTTCGCGACCGCGCGACACCGCTTTCGCTGACAGACGCGCTGCC 19389
QY 715 CTGGAAGCGCGCTTTCGCGCTTTCCTTCTTCGCGAGACGACCGCGCGCGCTTTCGCGAGTAC 774
DB 19390 GGGGCGGAGCTGCGCGCGAGCTTTCGCGAGGCGAGCGCTTTCGCGCGCGCGCGCGCG 19449
QY 775 CTCACTACCGGCTCGCGCGAGCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 834
DB 19450 GCGAAGCGCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19506
QY 835 CCGACCGCGCGCGCGCGCTTTCCTTCACTACCTGAGAGAGCACTCGAAGCGCGCGCGCGCGCG 894
DB 19507 AGCGTGAGAGCGCGCTTTCGCGAGCTTTCGCGAGAGAGCACTTTCGCGCGCGCGCGCG 19563
QY 895 GTGCGCGCGCGAGACTGCGCTTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 954
DB 19564 GAGGTGCGCGAGCTTTCGCGCTTTCGCGCTTTCGCGCTTTCGCGCTTTCGCGCTTTCGCG 19623
QY 955 CTGAAGCGCGAGACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 1014
DB 19624 TTGAAGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 19683
QY 1015 CTCTTTCGCGCGCGCGCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1074
DB 19684 GTCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 19743
QY 1075 CTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1134
DB 19744 CTGAGTGAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19803
QY 1135 CTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1194
DB 19804 CTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19839
QY 1195 CCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1254
DB 19840 CCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19899
QY 1255 CTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1314
DB 19900 CTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19959
QY 1315 CTGAACCAATGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1374
DB 19960 CTGAACCAATGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20019
QY 1375 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1434
DB 20020 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20079
QY 1435 AGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1494
DB 20080 AGCGCTTCGCGAGGCGGTTCTGCGCATCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20139
QY 1495 ATCAAGGAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1554
DB 20140 ATCAAGGAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20199
QY 1555 CTGCGCGCGCGGAGAGAGCGCGCGCGGAGACTGATGATGATGATGATGATGATGATGATGAT 1614

```



Db 20200 CTGCGACCTTGCGAAGAAAGACCGCGCCGAGAACCTGATGATCGACCTGATCCGCGAC 20259  
Qy 1615 GACCTCAACACAGCGCTCTGCGGATCGGCTCCGCTCCACGCTGCGCGGCTCTTTCGAGGTGAG 1674  
Db 20260 GACCTGGGGGGGGTGGCCGCGAGGTGCGCTCGGTGCTGCGCCGACCCGCTGTTCCAGGTTCGAG 20319  
Qy 1675 ACCTACGCGCGCCCTGTCACGAGCTGTGTTCGACCATTCGGGGAGACGGCTGCGCGCCCGGACC 1734  
Db 20320 ACCTACGCGCGACCGCTGTCACGAGCTGTGTTCGACCATTCGGGGAGACGGCTGCGCGCGGAGCAGC 20379  
Qy 1735 AGCAGCGCGCGCTGTCGCGCGCGCGCGCTTCCCGCGCGCGCTTCACGACCGCGCGCGCCAG 1794  
Db 20380 AGCGCGGTGGCGCGCGCGCGCGCGCTTCCCGCGCGCGATCGATGACCGCGCGCGCGAG 20439  
Qy 1795 AAGCGCACCATGAGATCATCGACCGCGCTGAGAGAGCGCGCGCGCGCTCTACTCCGAG 1854  
Db 20440 ATCCGACCATTCGACGATCATCGACCGCGCTGAGAGCGCGCGCGCGCTGCTCTCTCGAGC 20499  
Qy 1855 GCGCTCGGATGTTTCCGCTTCCGCGCGCGCGCGCGCGCTGACATCGCTCATCGCACCATTC 1914  
Db 20500 GCCATCGGCTACTTCTTCTCCCTCACCGCGCGCGGTGAGACTGTTCATCGTATCCGACGAGTG 20559  
Qy 1915 GTCCTGCGCGACCGCGCGCGCGCGCGCTGCGCTGCGCGCGCGCGATCGTCTCTCTCGAC 1974  
Db 20560 GTCTCAAGCGCGCGCGCGCGCTGCGCTACGCGCGCGCGCGCGCGCTCATCGCGCTCTCGAC 20619  
Qy 1975 CAGAGAGAGAGATTTCACCGACCGGTGTAAGAGCGCGCGCGCGATCGTCAACCGCGCTCGAC 2034  
Db 20620 CCGCGCGACGAGATTTCAGAGAGACCGCGCGGTGAGCGCGCGCGCGCTGCTGCTCTCGAC 20679  
Qy 2035 GCGAGCGCGCGTGG 2047  
Db 20680 ACCGCTTCCCGG 20692

RESULT 11  
STH575648 29132 bp DNA linear BCT 23-JAN-2004  
LOCUS Streptomyces thioluteus aureothin biosynthesis gene cluster  
DEFINITION (aurBCEBFGHI genes), strain HKI-22.  
ACCESSION AJ575648.1 GI:41322765  
VERSION acyl-coA ligase; aurA gene; aurB gene; aurC gene; aurD gene; aurE gene; aurF gene; aurG gene; aurH gene; aurI gene; cytochrome P450; N-oxidase; O-methyl transferase; PABA synthase; polyketide synthase  
KEYWORDS type I; transcriptional regulator.  
SOURCE Streptomyces thioluteus  
ORGANISM Streptomyces thioluteus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 He, J. and Hertweck, C.  
AUTHORS Iteration as programmed event during polyketide assembly; molecular analysis of the aureothin biosynthesis gene cluster  
JOURNAL Chem. Biol. 10 (12), 1225-1232 (2003)  
PUBMED 14700630  
REFERENCE 2 (bases 1 to 29132)  
AUTHORS Hertweck, C.  
TITLE Direct Submission  
JOURNAL Submitted (07-JUL-2003) Hertweck C., Bioorganic Synthesis, Hans-Knoell-Institute, Beutenbergstr. 11a, Jena, 07745, GERMANY  
FEATURES  
source 1. .29132  
location/Qualifiers  
/organism="Streptomyces thioluteus"  
/mol\_type="genomic DNA"  
/strain="HKI-227"  
/db\_xref="taxon:66431"  
2154. .2969  
/gene="aurD"  
2154. .2969  
/gene="aurD"  
/function="transcriptional regulator"  
/codon\_start=1

/transl\_table=11  
/product="AurD protein"  
/protein\_id="CAE02599.1"  
/db\_xref="GI:41322766"  
/translation="MVNAVVKLPDLSFNILCPVAVSTEGEPLPGPARQAVLSALL  
LAPGVQVAGERLLEAIWPAEPGNVAVSHSVSHRLRQLEBDCPEWGNRLIRRPQ  
GYLLAEARVDAARPEEGILREGRQLEIRGASQVIRACLEIRGSPVLELDGYE  
PAVCEAARLEERLRLATLEALMEALMGHDYATLAELAASARLPPTREIRLMVA  
LCEVROAAVAVVNRIRKALOBELGVDPGRRLRLFAILLDESGRACLOS"  
complement (3029..4546)  
/gene="aurE"  
/gene="aurE"  
complement (3029..4546)  
/gene="aurE"  
/gene="aurE"  
/transl\_table=11  
/codon\_start=1  
/product="putative acyl-coA ligase"  
/protein\_id="CAE02600.1"  
/db\_xref="GI:41322767"  
/translation="MLVPLTADFLDRAEATAARTALVDEBHPAALPSTYGETL  
RVRVMOAGLDALGIDEGEVRVAVSPNSARILELLVAVTASGRICVPVVRLLPEIS  
YILROCDASVVPVDPDVEAVVIGSVVPRKFTUGBOTETVEMFVAPRPMSPDENAV  
ATLNTSGTSARPKGVALTHRSITVMAVTFVHARTWBGDVLHTLPTFHCGKVPF  
LIAAGAQAVLIRIDGAEILRRVQEHGVTLLFGAPAVDAVLQAAADPGEIIPGRGR  
VAVVCGAPVSAQLVSRILENELGMEFLQVYGLTETTLTFNRLPDADPDGGLTRAGA  
PALGVRLRTGERGEVILARSNMVLDSYMRDEQNEALLEDGFWRTGDVADGEGHIAL  
PDRVADVITVGGESVSVEEDCLQGHSLAVAVVIGVDEMGFTVAVVVPADGAR  
VPEEVITVYCKQQLARYKAPTSVDRDALPRISNKLQGRLEREPMAGHDIRIH"  
4858. .5868  
/gene="aurF"  
4858. .5868  
/gene="aurF"  
/codon\_start=1  
/transl\_table=11  
/product="putative N-oxidase"  
/protein\_id="CAE02601.1"  
/db\_xref="GI:41322768"  
/translation="MREQPHLATWMAAGWVEEBGIGSATLGRVLRAPRRAAVVK  
ADILDEMDVDTLVDPDPLEIVPFAHEPLFLAEHPHORVLTGKMWIGNERVITGO  
LIAEPAFDLVNMGVPRGSDPDLIRSNVOALVDSFHYMHMLAIDRPELRKISER  
POPELVYTRRLARVLADMEPEQMERDIAVLWGAVAEFCINALLALARDATQPMHSL  
ITTLHLRDETHAGSIVAEVRELVARNMEOQRALVRCPLALEAFEDLSALLLET  
NNAAGIRGAEEIYDILRSTAGTFLVRDPSGARKWEOLGLDAVDPDFPERDWSPT  
PR"  
5990. .11710  
/gene="aurA"  
5990. .11710  
/gene="aurA"  
/codon\_start=1  
/transl\_table=11  
/product="polyketide synthase type I"  
/protein\_id="CAE02602.1"  
/db\_xref="GI:41322769"  
/translation="WELHGGDRSLRERLSAREADRRRLDLLIRTHSATYDGRAR  
ALDARAFGEQGVGRGKSAGKLELSEATGALSAVGFDTTPPALAEHLGLLGE  
DTAGPAAAAAADADEPIAIVGCGRLPGVNSPEGLMEVLSSETAISEPFDRCMDG  
EYDPPDHIGTIVTHAGVFEAPDPAIDPFAISPEGAVTIDPDRILLLETTWIEG  
ARLDTSLSGRSTGYFVGLMYSEYCARIRNPPABRGVAVGMSVYSGRLATYFGR  
BGPATVDTACSSLVAMHLAESLTKGECSLAVAGATWSTFWGYLFEKROGLAP  
DGRARSFADAAAGSSMSGCVGVVLELSDARNRGRVLAVLRGAVNDGASNGLTAP  
PNRPGQQOQVIRQALVHAGLSTADVVVDVHAGGTGLGDIPEQALLATYGQGRPADP  
LWLGSLKSNIGITQAAAGSVGVIKVMVIRNGVLPRLSHITPTFPVMSGAVELDT  
BARPMPERGRPRAGVSRFGVGTNAHYILRQAEPRPPAPASDGDGPAPLPMVSG  
RGAALRAQAGRIARFLRAHPREVPADVAISLAATRALPDAVVVYAADRAVARVLA  
ALADGEGEVALGVANRRDLAFVFEQGSQWPGMAELMSTSVFRESIEACKSL  
AEYDWSLTKVLRGRSGAPTLERVDVVPALFAVWVSALMRSGVLEPAAVGHSSQG  
EIAAAHVAGALSLDADARIIVARSKALVPLRGAGGVAAVPAVDELELEWDSGS  
VAAVNGPRAVVYISGDPGIDEAIEAFTRAGRAKIPVDVAISGAEVIEIRTLLEET  
AGIARRESEVPYSTVNGEVPDVTLLIDEMVITNLRVYQGPRTYRLIIDOCHFEVE  
MSPHVILVYGOETLIDVVAAGVLTFFDSLRRNGSTALRLTLEGAHNGIPVDAAQV  
FDHAGRPVDTLPTAFQRRRYWLEBAPEADALAGDGDVHPLGLAAVSLPDGTL  
FTGLSARTHPLTDHGVADVLVPGTALVDLALAGQVGTGEVLELVLAHPLPPE  
ODTVVQALCAPDESGRSVTVHARSSEARSLADPWTLLHAGETTLARDGERSDLPQAV  
PPAGAEIELDCYRQLAKNGLHYGAARQGLKRWRLDGDVFEVLELPESGPAAGYG



```

LHPALFDALHAAAFVGLPGAADAVATRLPFSMSGVALHATGAVRLARIRFTGPE
LSVTADPTGRPVVTTIASLGMRYTALHRAAASAGTSLIRPMPORAPADRALDE
ERVWITNGARGGOFERHDEPAGIRALADGATAAPVYIAELPMAADADATTCGA
HRVVRVLAQVOMSADRLAASRLVLTIDAVATASDAVAPVLAANWGLISAQOE
NPERIVLVLDGDESLAALPSALDTEAQLAVRAGSLVLRRLARVTTDPDRCAVALR
TIDVTVLTGGLVGRVLRVRLVTAHGIKRVLVYVSRGSGTEGVAJFTALNAGEAL
HVAAGADADAVLEGALASIPAEHPLTAVVMAGVLDGVLSTLTPERRARVLRPKAD
AAWHHELTAGDQVPLVYFSSLSVYGAGGGAANAAMDALVERBRAGSRGVAL
GMLKMGARSGLTGDDLEADLRMARVTVQPLSSDQALFLDLARAGDPVPVPLHDT
ALGCGDADVPVYLRGLARTPARARVRAEERPADDTLALRLAVPAADEAQLT
TLVRSHVAALVGYDDPTVEGRRTFTDIDPSLRALQLRNLKAGATGLRLPATVDEH
PTVEELGRYLRTRLVDPAPSAPTAEVTPPAAPGEGNAMEQLRSAAISLEVFLL
DQDLGK"
gene      11745..13832
CDS       /gene="aarg"
          /gene="aarg"
          /codon_start=1
          /transl_table=11
          /product="PABA synthase"
          /protein_id="CAE02603.1"
          /db_xref="GI:41322770"
          /translation="MRYLLIDNYDXXTYNLFOLLAEVYKPEVTLVNDSPWELDPA
MPCVVVSPGPRPOHPADLIGCOEVLDRWTDTPVLGVCIHQOALAHHGADVAEGTP
KHGHLARVHGRGDLFDGLPQNFVAVYHSLRAEPLPELBATNAMEDGTWALNR
TLPRMGVQHPRESIASEMGNOLONFPAKIVGAEYVAAPREBARMAPIEDPAEABL
RYEYVIDRAVDVTAELFTTLYKASAGCFMLDSSAPGIGPARFSLGDDSPRLAETLY
RTGENTVAANDAAGERAEPGTIFEVLQQRRLDRRLSTPELPLNGYGVYFYEELKG
ECGSPVNRSETPDVAWTFADRLIADVHLEKTYVAVSGTGEBAEAAMFETLR
RAELDPDPASEAAPAPATAPSEQVDIARQGVSDIKCDLHSGEAEYICLNIR
IRLPELKDPLGYLILRLNLPAPYGAFLRHGVSVCSSPERFLRIDRQVNSRPK
GTAPRGEDPADALRVSLAASAKTAENMIVDLRNDGRCCEVGVSVVAPYMTTE
TYATVHQLVTTVRGRLEHVDVACVACRACPGSGMGAFLRTMAIIDRLQGRGY
SGTYTGFGLSGAADLVIRTAVSTGTIVIGTGAIVLDSDEDEEVLKGLHALV
RAHGLEGAEA"
gene      13938..15158
CDS       /gene="aarg"
          /gene="aarg"
          /codon_start=1
          /transl_table=11
          /product="putative cytochrome P450"
          /protein_id="CAE02604.1"
          /db_xref="GI:41322771"
          /translation="WSTTAHTBPSWADLPFLDTPDNFMSDEPVAEARKNTATP
LALLVIRYAADQLARDKRLISGRGLVMDVNGTPEGVRFMVDFLOSLDGADHRLR
GLATHTPPTPRITAVOPFVASTVQOLIDLKPGDFPVQHPFPLPALVMCOLLGPL
EDYDTVGRSLSEITNLGLASNDODILVYKVOGGRMVDYVAALIKRKVYKGBDLSD
IVRAPHDGVLDDELRITLAVTVVAGVETTNHQLALAMDVFAHPIRPOOMKIKENPBLA
POAVERVLRMSPTLPVTATRVAAEDPEVNGVRLPTGTTPVMCAHVADRPRVADNR
PDTIVKREAPSIANGGPPHCLGTALARLELTAVALALATRLDPQIAGEITWRHEIG
VAGPDLALPLRFGAA"
gene      15244..21738
CDS       /gene="aarg"
          /gene="aarg"
          /codon_start=1
          /transl_table=11
          /product="polyketide synthase type I"
          /protein_id="CAE02605.1"
          /db_xref="GI:41322772"
          /translation="MTNDAKTLEYUKRLTALEETRELRITAEADQEPYAVVSMGR
YFGGVSSPEDMLVTDGTDIAAPFPADRGMVNDULFDPDPDRPGRTYTLLEGFVGA
AEFDADLFGISPREATAMDQORLLLETAEFERAGTDPGSLRGPVGFVGLSVGA
GSGGVGAEBAGBYHMTGNAASVLSGLAFAFLGPAVTVDAASALVAHQAVOAL
LRGECALALAGSIVTTTGTGFTERSROGLAPDGRCKAFATADGTTGEGGVAVL
LEKLSARKNKHGVLAIVIRGSANODGASGILTAPNGSPORVIRALAAARVSABEV
DIVBAHGTGTGLDPIVBAQALLATYTGGRDDPRLMIGSITKSLGHTTQGAAGVAGLIK
MVAHVHGVLPMTLVDEPSAHVMDSGAVRLLTGNHMDPETGRPRAGVSGVIGT
NAHLILEQAPDAESDAEPASGAPARIPWLARAGBALAEERAPVNGHTVAGLAEFL
VDVGHALTSRAALDQRAVVMADRGDLALATLAEBERAPVNGHTVAGLAEFL
SGGSGORPMGHETESFPVFAEKLEDCVGLHDLRDRPRELLFAEGREALAELO
TGYTOALFAHVEVALHHLTHMGIITPDLILGHISIGELTAAHVAGVLSLEBACALVAA
GRIMQOLPGAGAMLSVOATBAEVLPMVTEAHHEMSIAVNGPRSVVVSAGASVLEFA
EHKNGGRKTRKLRVSHAFSHSPOMDOMLOEPARVAKLAHPERRIVSVNSGEVATA
EOLCSFAYVVRAREVVRTHDTRILVABAGAFYFLEVGSGVCTTAAADLTABEPTV
TAAVSAGVPEADAAALAAVAEAVHVRWDMDFLEVGAGARRIDLTPTVARRSPFWI
OAPDADVTTAGLAGGHPDLGASLELADAOAGALSGRSARRESMLADHVAGSTLY
PGTAVEMVVRAGAEGBRLELTDEAPLAVERCAVILQVVGAGEGCHPVGY
SREDEPPEPMAGHARGVLAAPAAVPMPLGEGMGPVSGPRLPGFRVLAEGSA
YGRAPGGLRARRMLGEVLAELTLPREGACSGADRXYVNRALDAAHTLAKESANT
SOVRITPFAHRESFHGSGAPVLARLTPESGIDTVSLAMDEHETPVASGVLSVRPV
ARQLRATRHDTLFLDMWETITTPAAACAVIGDDELGAARVPFADIALVESADP
VPELVYPCIGDAEDDRADAASTLAVYVQAVADERMAVTTTLAVTRAMSATP
DREOVTDLVAAGVGLVRSQAQHPGRFPLAIDLDGDTASAALAPGLIASGDPQPL
REGAVLVPLRAGVPGSTLVPPGRTDMLELGGMGGVTDIALTPPEBAAPPLAPGV
RVAVRAGINFRDVWALGMVDRRALGGEIAGIYVAPRGVYGRAPGRVGLAOC
IGPVAVDRLRLARITPEGMSFPDAAVPTPLTAYGLVDLACVGRDRLVHAAAG
VGMAAVOLARHLRGAEVFATGAPAKMDVATLGDIDHLLASRTDEETRPAAADGGG
IDVANSIAGEMADSLRLVRPGGRFTEMGKTDIRADEAVAAVYEGVVRAPFLMDGG
AECIARIFAELTALFEGKIGLVPTVTTMVRQAPFAFRYFAQARHVKITLVTPPAMD
```



Db 12462 TCCGCGAGCCGCGGCTCTTCTGGCTGACAGACGCGCGCGGGATCGGCGCGCCG 12521  
QY 733 TTCTCTTCTCTCGGCGACGACCGGCGCCGCTCGCCAGTACCTACCTACCGCGTGC 792  
Db 12522 TTCTGTTCTCTCGGCGACGACTCCGGCCCTCGCGGACGCGTGAAGTACCGGACGG 12581  
QY 793 GACGGCTCTCTTCCTGCTCCGCGGCTCCGACGGACACGACCGGAGCGCGCCCTTC 852  
Db 12582 GAGAACACGGGTGCGGCGGACGCGCGCGGAGACGGGCGCGGCGCACCTTC 12641  
QY 853 TTCACTACTGAGAGACGAGTCCGAAAGCGCGAGCGGCTCCCGCGCCCGAACTGCC 912  
Db 12642 GAGTGTCTCAGACAGCGGCTCCGCGACCGCG-----GTTCAAGACGCGGAGCTGCC 12695  
QY 913 TTGAGTTCAACCTCGGCTACGTGCGCTACCTCGGCTACGAGTGAAGGCGAGACACC 972  
Db 12696 TTGACCTCAACGCGGCTACGTGCGCTACTTTCGGCTACGAGGCTCAAGGGCGAGTGCG 12755  
QY 973 GCGACCCCGCGACCGGTCCCGGACCCCGACCGCGCGTTCCTTCCTGCGGACCGCGC 1032  
Db 12756 AGCCCGAACCTGACAGCGGTCCGAAACCCCGAGCGCGCTTGAAGTTTCGCGACCGCTG 12815  
QY 1033 ATGCGCTCGACCCAGACGAAAGGTGCTGCTACCTGCTGCGCTTCGACCGC-----CG 1086  
Db 12816 ATGCGGTGACCACTGGAAGAGAGACCTTACGTGCGCTGAGCGGACCGGCGCG 12875  
QY 1087 GGCACGACGACGCGCGCGCTGCTGCGGAGACGCGCGAGACCTTACCGGCTG 1146  
Db 12876 GCGAGGAGGACGCGCGCGCTGCTGAGAGACCGCGCGCGGCGCGCGGACG 12935  
QY 1147 GCGGTCCGCGCGCGCGCGGCGGACCGCGGCTTCGGGATCCCGAGGCGCG 1206  
Db 12936 CCGGACCCCGCGCGCTGCGGACCGCGCG-----GCGCGCGACGCGCGCG 12983  
QY 1207 GCGGCTTCGCGCGCGCGCGCGCGCGCGCACGACAGGACGCGCTCAAGCGCATC 1266  
Db 12984 TCCGAACAGTTCACATCGCCCGGACCG-----GCGCATGCTGAGCGCATC 13034  
QY 1267 GACGAGTCTCAAGAGATCCGCAACGCGAGTCTGACGAGTCTGCTGACCAATG 1326  
Db 13035 AAGGACTGCTCGACGAGTGCACAGCGGAGGAGAGTACGAGTCTGCTGACCAACG 13094  
QY 1327 GTCAACGCGCGGACCGAGCGCGCTGCGCTTACTCTCGCGCTGCGCGCATGAC 1386  
Db 13095 ATCCGCTGCGGACGCTGAAGACCGCGTCCGCTACCTCTGCTCCGCGCTCAAC 13154  
QY 1387 CCGTCCCGTACGCGCGCTGCTGAGTTCCCGAACTGTCGAGTCTGAGCGCTGCGCC 1446  
Db 13155 CCGCGCTTACGCGCGCTTCTGCGCACGAGGCGTGAAGCTGCTGCTCTCCCG 13214  
QY 1447 GACGCTTCTCAAGTCCGCGCGGAGCGCGGCTCGAGTCAAGCCCATCAAGGGAGC 1506  
Db 13215 GACGCTTCTGCGGATCGACCGGAGCGGCTGTGAGAGCGCGCATCAAGGGAGC 13274  
QY 1507 CCGCGCGCGCGGACCGCGGAGAGAGAGAGCGGCTCCGCGCGCATGCGCGCG 1566  
Db 13275 GCGCGCGCGGCGAGGACCGCGCGCGGAGAGAGCGCTCCGCGCTGCGCGCTCG 13334  
QY 1567 GAGAGGACCGGCGCGGACGATGATGTCGACTGTCGCGCAACGACCTCAAGC 1626  
Db 13335 GCGAAGACCGGCGCGGACCTCATGATGTCGACTGCTCCGCAACGCTGCGCGG 13394  
QY 1627 GTCTGCGCATCGCTCTGTCAGTGCCTCCGCTTTCGAGGTGAGACTTACGCGCC 1686  
Db 13395 GTGTGGAGGTTCGTTCTGTCAGTCCCGCTCATGTCACCGAGAGTACGCGAGC 13454  
QY 1687 GTGACACAGTGTGTCATTCGCGGAGCGCTGCGCGCGCGCACCAAGACCGCGCC 1746  
Db 13455 GTGACACAGTGTGTCACAGTTCGCGGCGCGCTGCGGACGAGTGCAGCGCGCTCG 13514  
QY 1747 TGGGTACGCGCGCGCTTCGCGCGCGCTCATGACCGCGCGCGCGCAAGAGCGCACATG 1806  
Db 13515 TGGGTCCGCGCGCTTCCTCCCGCGCTCATGACCGCGCGCGCGCAAGCTGCGGACGATG 13574

QY 1807 GAGATCATGACCGCTGAGAGAGCCCGGCGGCTACTCCGGGCGCTCGAGTGG 1866  
Db 13575 GCCATCATGACCGCTGAGAGAGAGCCCGGCGGATCATCTCCGGCACATCGGCTAC 13634  
QY 1867 TTGCGCTTCAAGCGCGCGCGGACCTTACGATCTGATTCGACCATCTGCTGCGCGAC 1926  
Db 13635 TTGCGGCTTCCGCGCGCGGACCTCAACATCTGATCCGACCGCGCTTCACCGGT 13694  
QY 1927 GCGCAGCGGAGTTTCGCGCTGCGCGGCGGAGATGTTGCTTCGACGAGAGAGAG 1986  
Db 13695 ACGGCGACGCTATCGGCGACCGCGCGCGCATGCTTCTGACTCGGACCGGACGAGAG 13754  
QY 1987 TTCAACGAGACCGTGTAAAGCGCGCATGTCACCGCTCG 2032  
Db 13755 TTGACGAGTGTCTCAAGGCGCACGCGCTGTGTCGCGCGGACG 13800

RESULT 12  
AP00618 17/c  
WPCOMMENT

Sequence split into 61 fragments	LOCUS	AP00618	Accession	AP00618
Fragment Name	Begin	End		
AP00618_00	1	110000		
AP00618_01	10001	210000		
AP00618_02	20001	310000		
AP00618_03	30001	410000		
AP00618_04	40001	510000		
AP00618_05	50001	610000		
AP00618_06	60001	710000		
AP00618_07	70001	810000		
AP00618_08	80001	910000		
AP00618_09	90001	1010000		
AP00618_10	1000001	1110000		
AP00618_11	1100001	1210000		
AP00618_12	1200001	1310000		
AP00618_13	1300001	1410000		
AP00618_14	1400001	1510000		
AP00618_15	1500001	1610000		
AP00618_16	1600001	1710000		
AP00618_17	1700001	1810000		
AP00618_18	1800001	1910000		
AP00618_19	1900001	2010000		
AP00618_20	2000001	2110000		
AP00618_21	2100001	2210000		
AP00618_22	2200001	2310000		
AP00618_23	2300001	2410000		
AP00618_24	2400001	2510000		
AP00618_25	2500001	2610000		
AP00618_26	2600001	2710000		
AP00618_27	2700001	2810000		
AP00618_28	2800001	2910000		
AP00618_29	2900001	3010000		
AP00618_30	3000001	3110000		
AP00618_31	3100001	3210000		
AP00618_32	3200001	3310000		
AP00618_33	3300001	3410000		
AP00618_34	3400001	3510000		
AP00618_35	3500001	3610000		
AP00618_36	3600001	3710000		
AP00618_37	3700001	3810000		
AP00618_38	3800001	3910000		
AP00618_39	3900001	4010000		
AP00618_40	4000001	4110000		
AP00618_41	4100001	4210000		
AP00618_42	4200001	4310000		
AP00618_43	4300001	4410000		
AP00618_44	4400001	4510000		
AP00618_45	4500001	4610000		
AP00618_46	4600001	4710000		
AP00618_47	4700001	4810000		
AP00618_48	4800001	4910000		
AP00618_49	4900001	5010000		



AP006618_50	5000001	5110000
AP006618_51	5100001	5210000
AP006618_52	5200001	5310000
AP006618_53	5300001	5410000
AP006618_54	5400001	5510000
AP006618_55	5500001	5610000
AP006618_56	5600001	5710000
AP006618_57	5700001	5810000
AP006618_58	5800001	5910000
AP006618_59	5900001	6010000
AP006618_60	6000001	6021225

Query Match	32.2%	Score 662.8	DB 1	Length 110000
Best Local Similarity	60.7%	Pred. No. 4.5e-46		
Matches 1269	0	Mismatches 742	Indels 81	Gaps 8

OY	1	TTGGGCAAGCTTTGATGATGCAACATTAAGACTGCTTACCCCAACAACCTGTTCCAGTATATC	60
Db	54159	ATTGGGCAAGCTGTTGATGATGCAACATTAAGACTGTTTCACTCAACAACCTGTTACAGCTGATC	541000
OY	61	GGCGAGGCAACCGGAGCAACCCCGTGTGTGTGCGCAACGACG---CGACTGTGTGCGG	117
Db	54099	AGTAGAGGTAAACGGGGGTGCAACCGACCGTGTGTGCGCAACGACGACCGACCGGCGCG	540400
OY	118	CTGCGCCGTTCAAGGACTTTCGACGGCATGTGTGTCTCCCGGAGCCCGGCAAGCCCGAACCGG	177
Db	54039	CTGGAACCTGCGCGCTTACGACAAACATGTGTGTCTCCCGGAGCCCGGCGCGCGCGGACCGG	539800
OY	178	GAACGGGACTTTCGAAATACACCGCGCGGCGCATACGACAGCGGCGTGCCTGCTTCGCG	237
Db	53979	ACCGGCGACGTCTGGGGCTGTGTGCGACGGGTATGACGCGGCGCTCCAGCGTGTGCTCGCG	539200
OY	238	GTCTGCTCTCGGCAACGAGGATGCGCCAGCTCTTGTGGCGAAACGTGTGACCTTGTGCGCGG	297
Db	53919	GTGTGCTCTGGGGCAACGAGGATGTGTGTGTGCGGCGGCGAGAGGGTTCAGACCGCGCACCG	538600
OY	298	GAACCCATGCAACGGCGCGGGGTCTCCGAGTGTGGGACACACCGGCGAGGACGCTTTCCGGGCG	357
Db	53859	GTACCCCGGACGCGCTTCTCGACGGGTGTGGGACACGACGGGCGGACCTGTTCGCGCGG	538000
OY	358	CTCCCTTCGCGCTTACACCGCGCGTGTGCGCTACACACTCCCTGTGCGCGCACCGA---CCTGCC	414
Db	53799	CTGCGCGAGGATTTCAAGGTGTGTGGGTATCACTGTGTGTGTGGCGGTGTGAAGCGCTGTGCGG	537400
OY	415	GACGAGCTTCGAACCCCTTCGCTGTGAGCGACGACGAGGGTGTCTCATGSGGCTTGTGGCAACCGC	474
Db	53739	CCCGCCCTCGAGGTGACTGTGCGCGCACCGCGAGGGGTGTATCTGTGGGGGTGTGGGCAATCGG	536800
OY	475	GAGAAAGCGCGTGTGGGGCGTGCAGTTTCAACCGGAGTGCATGTGGGAGCGCACTTTCGGCGCG	534
Db	53679	CAGCGCGCGGAGTGGGGGTGTGCAGTTTCAATCCGAGTGTGTGTGGCGCGGCGAGTTCGGGCGG	536200
OY	535	GAGATCATGTGCAACTTTCGCGCACTTCGCGCTTCGCGCAC---CACCGGAGACGGCGCG	589
Db	53619	GGCGTGTGTGGGCAACTTTCGCGGAACTGTACCTGTGGCGGACGGGAAACCGCGCGGAGAGAG	535800
OY	590	ACGGGCGCGACTCCCGGTAAGAACTCCAAGTGTGCGCGCGGTGCAGCTGTGTCGGAGCGCG	649
Db	53589	CGTCCGGCGGACCCCGTGAAGCGCGCGCGCGTGTCCGAGCGCGACCGACCGGCGGTGTGGCGG	535000
OY	650	AAGAGGTATCGCGCGGCGGTGCGCTGCGCGG-----	677
Db	53499	CTGGCGGACGAGGTGATGTGAGCGCGCGGTGTGACACCGAGGCGCGGCTTCTCGCGGCTTAC	534400
OY	678	-----CGAGGCGACCAAGTTCGTGTGTGACAGAGCTCGTCTTCGAGAGCGCTTCGCGG	732
Db	53439	GGCGCGTGTGCGAGCGGGGTTCGTGTGTGACAGAGAAACAGTGTGAGCGGGGCTGTGACCGG	533800
OY	733	TTTCTCTTCTCTCGGCGAGACCGGGGCGGCTGTGCGAGTACTCTCACTCAACCGCGTGTGCG	792
Db	53379	TTTCTCTTCTCTCGGGTGACGAGAGGACCGCTGTGCGAGAGGTGTGTGTGCTTACCGGGTGTGT	533200

QY	793	AACGGCGTCGTCCTCCGTCGCGGCTCGAAGGACACAGACCCGAGCGGGCGCCCTTC	852
Db	53319	GCGGACGTCGTGGCGGGCTGAGGACGCGCGGGAAACCGGACGGTCCGCGGGACGTGCTG	53260
QY	853	TTCAACTACCTGGAGGAGCAGCTCGAACGCGCGAGGATCCCGTCGCCCGCAACTGCC	912
Db	53259	GGTTATCTGTCCACCCACAGCTGCGCGA-----CGGACGTGAGAGTGTCCCGACATGCCG	53206
QY	913	TTGAGATTCAACCTCGGCTACGTCGGCTACCTCGGCTACGAGCTGAGAGGCGGAGCAC	972
Db	53205	TTTCAGATTCTCGCGCGGATAGTCGGGTACTCGGCTACGAGGTCAAGGCGCAGCTCGCG	53146
QY	973	GGGCAACCCCGCGACCGGTCCCCGACCCCGAGCGCGCTTCTCTTCGCCGACCGCGCC	1032
Db	53145	GCGGAGGCGCGGACCGGGCGCGCACCGAGCGCCAGTGTCTTCGCCGACCGGCTG	53086
QY	1093	ATGCGCCCTCGACCAACGAGAGGCTGTGCTACTCTGTGCGCCCTCGACCGCGGGGCGAC	1092
Db	53085	ATCGTGGTCAACCACTGTGCGGGGCAAGACCACTGTGTGCGCTCAC-----GACGAC	53032
QY	1093	GACGACGCGCGCGCGCTGCTGCGGGAGACCGGCGGAGACCTTCACCGGCTGCGCTG	1152
Db	53091	GACACCGCGGCGCATAGTGGCTGCGCGAG-----ACGGGGGGGTGCTG	52987
QY	1153	CGCGCCCGCGCGACCGCCGATGTGTTTGAGATCCCGAGCGCGCGCGCG	1212
Db	52986	GAAACGCTGCGCGGTGGGAGATCCGCGGAAATTGGCATGAGTCCGCGCGGCTTG	52927
QY	1213	TTGGGCGCCCGCGCGCGCGCGCGACGACAGAGAGCGCTACTCTCAAGCGCATCGACGAG	1272
Db	52926	GTCGCGCGCTGCTGAACCGGGGCGGACGG-----TACTGCGGACATCGCGCG	52873
QY	1273	TGCTCTCAAGAGATTCGCGACCGCGAGTGTGTCGAGATCTGCTTGACCAACTGTGTACC	1332
Db	52872	TGCTCTGAGATCTGCGCGCGCGCGGAGTCTTGAGATCTGCTTGACGACAGCTCAAC	52813
QY	1333	GCGCGGACCGAGGCGACGCGCTTGCGCGCTTACTCCGCGTGGCGGCATCGACCCCGTC	1392
Db	52812	GTCGACCGCGCGATCGGCGGCTGAGCTTCTTATCGACCGCTGCGCGCTGCAATCGGCG	52753
QY	1393	CCGTAAGCGCGCGCTCGAGTCCCGGAACTGTGGTGTAGAGCGCTCGGCGGAGCGG	1452
Db	52752	CCCTATCGCGCTACGCTGCGCTTTCAGACCTCGAGATGCGCTGCTCTTCACCGGAGCGC	52693
QY	1453	TTCTCTCAAGATCGGCGCGCGACGCGGCGTCTGAGTCCAAACCCATCAAGGGGACCGCGCC	1512
Db	52692	TTCTCTCAAGATCGACCGGGGCGCGACGCTGAGAGGAAACGATCAAGGGACCGCGCG	52633
QY	1513	CGGGGCGGCAACCGGAGAGAGCGAGCGCTTCGCGCGACCTGCGCGGCGGAGAG	1572
Db	52632	CGCGGGGCGACCCCGCGGAGGACGAAACGGCTGCGCGGAGGTTGGCGCAGCTGCCAAG	52573
QY	1573	GACCGGGCGGAAACCTGATGATCTGTTCGACTGTGTCGGAAGACTTCAACAGCGTTCG	1632
Db	52572	ACCGGCGCGGAAACCTGATGATCTGTTCGACTGTGTCGGAAGACTTTCGAGGCGTTCG	52513
QY	1693	GCGATTCGAGCTCCGTCACCGTGCACCGCGGCTCTTTCGAGGTGAGAGCTCAACCGCGCGTGC	1692
Db	52512	CAGATTCGAGCGCTGACAGTACCGCGAGCTGATGAGCGAGAGTCTGATTTCACCGTTCAC	52453
QY	1693	CAGCTGTGTTCGACATTCGGGAGCGAGCTGCGGCGCGGACCGACGACCGCGCGCTGCTA	1752
Db	52452	CAGCTGTGTTCGACGAGGCGCGGACCCGTGCGCGCGAGCTGCGAGCGCCATCGACTGCGTG	52393
QY	1753	CGCGCGCGCTTCCCGCGGCTTCATGACCGCGCGCCCAAGAACGCACTGAGAGTTC	1812
Db	52392	CGGCGCTGTTTCCCGCGGCTTCATGACCGGCGCGCCCAAGCTGCGGACGATGAGATTC	52333
QY	1813	ATCGACCGCTTGGAGGAGCGCCCGCGGCGTCTACTCGGGGCGCTCGGATGTTCTCC	1872
Db	52332	ATCGACGAGCTGGAAGCGCGCGCGGGGTCTACTCGGACCAATCGGCTTCTTCGGA	52273
QY	1873	CTCAGCGGCGCGCGGACCTCAGCATGTCTATCCGACCATGCTGTGGCGGACGCGCAG	1932



```

Db      52272 CTGGCGGCGCACCGCCGCACTCAATGTCATCCGACCGCGGTGCGCACGACGGCGC 52213
Qy      1933 GCCGAGTTTCGCGCTCGAGCGGCGGCGATGTCCTCCGACCAAGAGAGAGATTACAC 1992
Db      52212 TGCGCGATTCGGCGCGCGCGCGCGATGTCGTGATTCGATCCGATCCGCGGAGAGATATCCG 52153
Qy      1993 GAGACCGGTGTAAAGGCCCGCCGATGTCATCCGCCCTCGACGCGACGCGC 2044
Db      52152 GAGATGTTGTGAAGAGCGCGCGACCCAGCGCGCGCCACCGCGCGCGC 52101

RESULT 13
LOCUS   AP005025/c
DEFINITION Streptomyces avermitilis genomic DNA, complete genome, section 5/30.
ACCESSION AP005025 298300 bp DNA linear BCT 10-May-2003
VERSION   AP005025.1 GI:29604584
KEYWORDS  Streptomyces avermitilis MA-4680
SOURCE    Bacterial Actinobacteria; Actinobacteridae; Actinomycetales;
ORGANISM  Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE
AUTHORS  Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
          Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
          Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
          Genome sequence of an industrial microorganism Streptomyces
          avermitilis: deducing the ability of producing secondary
          metabolites
          Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
          21477403
          11572948

TITLE     Ikeda,H., Ishikawa,J., Hanamoto,A., Shinose,M., Kikuchi,H.,
          Shiba,T., Sakaki,Y., Hattori,M. and Omura,S.
          Complete genome sequence and comparative analysis of the industrial
          microorganism Streptomyces avermitilis
          Nat. Biotechnol. 21 (5), 526-531 (2003)
          22608306
          12692562

JOURNAL   MEDLINE
PUBMED    3 (bases 1 to 298300)
REFERENCE  Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
          Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
          Kushida,N., Director-General of Biotechnology Center, Shiba,T.,
          Sakaki,Y. and Hattori,M.
          Direct Submission
          Submitted (29-MAR-2002) Director-General of Biotechnology Center,
          National Institute of Technology and Evaluation, Biotechnology
          Center/ 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
          (E-mail:bio@nite.go.jp, URL:http://www.bio.nite.go.jp/,
          Tel:81-3-3481-1933, Fax:81-3-3481-8424)

TITLE     This work was done in collaboration with Haruo Ikeda(*1), Jun
          Ishikawa(*2), Akiharu Hanamoto(*3), Chigusa Takahashi(*3), Mayumi
          Shinose(*3), Hiroshi Horikawa(*4), Hidekazu Nakazawa(*4), Tomomi
          Osonoe(*4), Norihiro Kushida(*4), Hisashi Kikuchi(*4), Tadayoshi
          Shiba(*5), Yoshiyuki Sakaki(*6,*7), Masahira Hattori(*1,*7)
          and Satoshi Omura(*1,*3).
          Final finishing process and all annotation were done by H. Ikeda
          and J. Ishikawa.

COMMENT   *1 Kitasato Institute for Life Sciences, Kitasato University
          *2 National Institute of Infectious Diseases
          *3 The Kitasato Institute
          *4 National Institute of Technology and Evaluation
          *5 School of Science, Kitasato University
          *6 Institute of Medical Science, University of Tokyo
          *7 RIKEN, Genomic Sciences Center
          Following url is also available.
          http://avermitilis.nis.kitasato-u.ac.jp.
          location/Qualifiers
          1. .298300
          /organism="Streptomyces avermitilis MA-4680"

```

```

gene     /mol_type="genomic DNA"
          /strain="MA-4680"
          /db_xref="taxon:227892"
          /note="This strain is also named as strain: ATCC 31267,
          NCIMB 12804 or NRRL 8165.-synonym: Streptomyces
          avermitilinus"
          208. .921
          /gene="orf-1"
          /gene="orf-1"
          /note="SAV944
          not involved in biosynthesis of avermectin"
          /codon_start=1
          /transl_table=11
          /product="reductase"
          /protein_id="BAC68654.1"
          /db_xref="GI:29604585"

          /translation="MCVDRDGRGAGVTDLADLTARQAAPMEVADVSDGAMERPAER
          VAEITYGVDLVNMAGIGMAGRFDDTSVEDMORTIGVIMAGVHIGRLIGRQMARQ
          GGHIVTVASAAAFQPTPAVPAVATSKAAVIMSECLRAEFAGFVGVSVCQFVRS
          PASAMHFAGVPRLEOERLRALFPAAGCSAEKVAAVLRVARDAVVVTAAERLSRL
          MRFTRPLRAAARMDPPS"
          979. .2217
          /gene="aveB1"
          /gene="aveB1"
          /note="SAV945
          avermectin gene cluster"
          /codon_start=1
          /transl_table=11
          /product="glycosyl transferase"
          /protein_id="BAC68655.1"
          /db_xref="GI:29604586"

          /translation="MSDHLFMSAPFMGVFPVPSLAVAEELVHGHVTFVTCGAADA
          VASGADPRKYSABGVDMYRLMTAEARNAPMTLYDEGSMLSVEHVKKDVPDL
          VADTATLSNVRVLAASWSRPAMTVIPFASNGRSTMQSVLDPDSAOVASAPREFS
          EOMELFGALAPRLAEELVSRGITEPVDDFLSGDEPNVLDPAPAFQVAGTFFERF
          AVGVCPLGRKRGIGMTFPGSGHPVLISLGVFROUSFFRTFVRAFTVDVPHVTS
          LKGVDPDLRPLPNNVEHVRVPHAVLEHARALVTGCTGSVMEALHAGCPVLMP
          LSRDQVGRRIAEILGIMVQPEEVTATTLARRHVLDIISDAITRQVRQMRATVEA
          GGALAAADBTERTFLRTRRH"
          complement (2123. .3190)
          /gene="aveB11"
          /complement (2123. .3190)
          /gene="aveB11"
          /note="SAV946
          avermectin gene cluster"
          /codon_start=1
          /transl_table=11
          /product="dtpd-glucose 4,6-dehydratase"
          /protein_id="BAC68656.1"
          /db_xref="GI:29604587"

          /translation="WTGAGRFIGSHFRRRLTGAYPATGAEVVVLDKLTVAERLENT
          AVLGSPSLIEVHGDI CDGPELVADLMDSDVNVHPAASHVDRVADAAEFVRTVLG
          THTLLRAADLVADRFRVYISTDEVGSI DSGWTDADLPENSPVSAKASLDLARS
          FRRTGKLPVYIRCSNNYGHQFPKLI PRFTYHLINGTQVLYDGEVTRMDLVSD
          HORQGLVAERGRPEGEIYHIGGTLNSRELTRALLDLGVMSWVEPVTDRKGDRR
          YSLDISAISAEIRGVAHYVPEBGGLAQTYQWYVENRTLMEPLTARELPEVSDASAEAT
          AASRPLPACGRPRFPWPAASA"
          complement (3187. .4086)
          /gene="aveB11I"
          /complement (3187. .4086)
          /gene="aveB11I"
          /note="SAV947
          avermectin gene cluster"
          /codon_start=1
          /transl_table=11
          /product="glucose-1-phosphate thymidyltransferase"
          /protein_id="BAC68657.1"
          /db_xref="GI:29604588"

          /translation="WKGIVLAGCTGSRLYPLTRALSQKLLPYDYKPMYIYPLSLMLG
          GIKDLIVSSPDHLQFRRLPDGSRIGLANTDYAAQQRPGGIAEFLIGADETGGDOV
          SILVDNITFHGWSFHLRSHRTVDGCVLRFQAVTDPBERYGVGEVSDASGKLLSVEEK

```







[illegible]

DB	265829	TCG	265827
RESULT 14	AP005217	300750 bp	DNA linear BCT 24-JUL-2003
LOCUS	Corynebacterium efficiens YS-314 DNA, complete genome, section 4/11		
DEFINITION	AP005217	BA000035	
ACCESSION	AP005217.1	GI:23492722	
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL	</		



```
gene
  /protein_id="BAC17696.1"
  /db_xref="GI:23492724"
  /translation="MSSQTPNSMPPTVHHHLDGVFADPWWIHPITTCWMPQPHGSNT
  NMPHEHSPWVVV"
  2399..5481
  /gene="rRNA_23S"
  2399..5481
  /gene="rRNA_23S"
  /product="23S ribosomal RNA"
  /note="CE-RNA02"
  5606..5725
  /gene="rRNA_5S"
  5606..5725
  /gene="rRNA_5S"
  /product="5S ribosomal RNA"
  /note="CE-RNA03"
  5925..7088
  /note="CE0889, similar to AX063829-1|CAC25156.1| percent
  identity: 79 in 387 aa"
  /codon_start=1
  /product="putative aminotransferase"
  /protein_id="BAC17699.1"
  /db_xref="GI:23492727"
  /translation="MKNKYVERLRPFGETIFATMTORANAGANINGQFPEDGPA
  PMLEIAQOILGNNQYSGRGLDRTAVASDHRKFKLDYDPTLITGVATBAI
  SASVLGVERGDEVILEPYDAAYAAIALAGATRVAVPKEEDNTMLDTDAVHAIV
  TKRTSMITVNSPHRPTGVSFSKSLMALIARAYDLIVSDEYHETLTDKTSHTAV
  ASLPGMDRTITVSSAKRTNVGKMGALAPVILEAVIRAKQVSVGATPQFA
  VAHVAHEKPMVKQKGLQNKDILIDALTOAGLKVDHGTGFVADIGEDGAEF
  CFPILDRGVAAIPVQAFVDRKEQMSKVFAPCKQEDTLREAAQRLRAAGSL"
  complement (7180..7821)
  /note="CE0890"
  /codon_start=1
  /translation="hypothetical protein"
  /protein_id="BAC17700.1"
  /db_xref="GI:23492728"
  /translation="MTHSFVDVAVAKKNEPLDTRKLIQISAFRGLVILAGVL
  YIADGAVKGMWLIYKGIKALSPFMIPIVLPPOVGSASLINTYPLVAVAEINRK
  DVVLLAVNTNVEETLPPRGVLAFTINRGVDRKLGEEVPSVAASGRTIIDDQHM
  DEISPMPTIGTIDADVRAEKTIPIDLWNKLEKNNRILDEVNTMNLVLV"
  complement (7822..9495)
  /note="CE0891, similar to AL583924-93|CAC31112.1| percent
  identity: 75 in 547 aa"
  /codon_start=1
  /translation="hypothetical protein"
  /protein_id="BAC17701.1"
  /db_xref="GI:23492729"
  /translation="MARGDPLIVOSDKYULLETHDLAGEARALAPFAELERAPH
  VHYRTITPLAMNARAGHDQVVDLEYSRPVQPLLVDAETWSYGVGRMKK
  HPAGLILLESSEPAIIVEISRHRIKPMLOQDPTETIAVPSERGLKQBLKIGAP
  AEDLAGVDESEPAIISTEVEDMSLEDYQOADSWEQSGGVVLPCCAGTMTVA
  ASNARQATLILVTNVAGROWKDELIRTLTLEIDEGYSGERKIRLVTLATYVD
  VTRRTGGEYKALEPDSRDWGLIIVDEYHLLPAPVPEMTSDLOSRRKLGTLTIVED
  GRGQVUSLIGPKRYDAPKMDLEAGCIATADCEYVTTMTDAERMYATAETADRYR
  LAATVHTKTVVRRILIEHQGPTLIIIGALVDQLEIGAEFDAPVYDGKTSNKKRGSL
  FDRFSGELNVLVSKVANSIDLPEAAVAIIVSGTFSRQEBEORLGRLLRPKADGP
  EAHFYVSRTDTEYAAHRORFLAEQGYAHRIIDADDLLPITEKET"
  complement (9548..11851)
  /note="CE0892, similar to AE006976-11|AAK45126.1| percent
  identity: 36 in 741 aa"
  /codon_start=1
  /translation="conserved hypothetical protein"
  /protein_id="BAC17702.1"
  /db_xref="GI:23492730"
  /translation="MPCFPSSQIRQTVSTARKLDGIMKMSBPVITLTWLEQDLD
  QLSGILNRPDITLPLPGLGSLAARQLQASIORAVHKLASLELALEANAAGL
  NPITAPVEVERLHRLAHSAAVPTAEQILNRLRLALVFGGERIILIPETMALTP
  TMMQLPEEHRSLITFGEASOSIALPARHKLIVTLVTNSGLTLDALSDAPAP
  IPIKIKGLSLRDEQTVKLPGVVRRVLLEGEDPTAMRVVPEBPSAVPGADAGIAGL
```

```
gene
  /protein_id="BAC17703.1"
  /db_xref="GI:23492731"
  /translation="MGNMANVEKKGYVDPAMPDHPDSEGHVTELIAPVAGASSPMWD
  DMEFPVPAERTIGVHPYTRINR"
  complement (12229..12822)
  /note="CE0894, similar to AX133781-1|CAC38957.1| percent
  identity: 71 in 195 aa"
  /codon_start=1
  /translation="conserved hypothetical protein"
  /protein_id="BAC17704.1"
  /db_xref="GI:23492732"
  /translation="MLPMGRHSAKKTITPPYKMAATTVARGAAVAAPAPASAPSDW
  LRAQESGGMACINTGNGHGLQSPSTWRAVGDDEFAFPYQATREQOIAVET
  LIAQGGGAPMACKSLGINSAPTRPMLNAPAPAPAPAPAEYRAVDTNTNVS
  DLNTVQIGIDAVQVLAQYGVQVPAEIQAHYNAFGR"
  13342..13725
  /note="CE0895, similar to AE006977-4|AAK45135.1| percent
  identity: 52 in 127 aa"
  /codon_start=1
  /translation="putative cold shock protein"
  /protein_id="BAC17705.1"
  /db_xref="GI:23492733"
  /translation="MPVGTWKYDPERGGEFVNSPGEDCFVKQVLPKATVLEHQO
  RLEFDPAGRGQALRAVVLPTPRRPHQTYPERLNGIISDMVTMLBGTVPALIK
  GHYDPRKRVQVKKIIRAVAKELAA"
  complement (13881..14408)
  /note="CE0896"
  /codon_start=1
  /translation="hypothetical protein"
  /protein_id="BAC17706.1"
  /db_xref="GI:23492734"
  /translation="MASRKTGRKQFIQATILVAVVVVVIVSVQNMNMNRPEPE
  TNSITSSPAGELIEVPYSICERGVCAENEVFTLVGADDELRLTIPBIIHRHDTL
  LTIYDVPANDEYHYSYETEVTVGSVDPVTEGGERPLVVVEVTSVMIGTDNGE
  EFPYTTWISLATVQ"
  complement (14424..15254)
  /note="CE0897, similar to AE001874-1|AAF09704.1| percent
  identity: 40 in 267 aa"
  /codon_start=1
  /translation="putative glutamine cyclotransferase"
  /protein_id="BAC17707.1"
  /db_xref="GI:23492735"
  /translation="MPCFPSSQIRQTVSTARKLDGIMKMSBPVITLTWLEQDLD
  QLSGILNRPDITLPLPGLGSLAARQLQASIORAVHKLASLELALEANAAGL
  NPITAPVEVERLHRLAHSAAVPTAEQILNRLRLALVFGGERIILIPETMALTP
  TMMQLPEEHRSLITFGEASOSIALPARHKLIVTLVTNSGLTLDALSDAPAP
  IPIKIKGLSLRDEQTVKLPGVVRRVLLEGEDPTAMRVVPEBPSAVPGADAGIAGL
```

```
Query Match      23.2% Score 478.4; DB 1; Length 300750;
Best Local Similarity 57.6%; Pred. No. 2.4e-31;
Matches 1164; Conservative 0; Mismatches 676; Indels 180; Gaps 10;
```



Db 185000 ACGAGACGCTGTTCAGCGCCGTATATCTCAACCCCGGCCACCCCGGGGTGCGG 185059

QY 182 GGGACTTCGGAATCAGCCCGCGGCATATCCAGACAGCGGCTCGCCGTCCTCGAGCTC 241

Db 185060 CGGATTTCCGACAGCTGCCCGGGGTGATCGAAGCTGTGTCGTCCGCTCTCGGGGTG 185119

QY 242 GCCTCGGCCACAGGGGCATCGCCCACTCTTCGGCGGAACCGTCGGCTTCGCCCGGAAC 301

Db 185120 GCTTGCGGCATCAGGGGCATCGCGCTCGCCACCGCGCGCGGTGGGCGCGGCGCGTCC 185179

QY 302 CCATGACGGCCGGGCTCTCGAGTTCGGGACACCGGCGAGAGCGTCTTCGGGGCTCC 361

Db 185180 CGGTGACGGTCAAGGTACCCCGCATCACACACGAGTCCGAGCTTTTCAGCGCATCC 185239

QY 362 CCTCGCCGTTCAACCGCGGTGCGGTACACTCCCTGCGCCGACACGACTTCCCGAGAC 421

Db 185240 CGCCCGAGTTTCAGCGCGCTTCGTTACACTCCCTGTGTGGCACCGAATCTGCGCCGGAAC 185299

QY 422 TCGAACCCTCTGCGCTGGAACGACGAGGGGTGTCATGAGGCTTCGCGACACCGAGAAAG 481

Db 185300 TCGAGGTTCACCGCGAGGACCGGGGACGGGCTGATCATGCGGTGCGCCACCGGAGCTGG 185359

QY 482 CGCTGTGGGGCGCTCCAGTTTCACCCCGAGTTCATCGGCACGCACTTCGCGCGGGAATCA 541

Db 185360 CCGAGTGGGGGTGCGAGTTTCACCCCGAATTCATCGCGGGCGGTTTCGGGCAACCGATCA 185419

QY 542 TGGCAACTTCCGCGCACTTCGCCCTCGCCACACACGGGGACGGGGCCACCGGGCGGACT 601

Db 185420 TGGCGAATTTCTTGAGCCTTGGC-----GCTTGACACAGCGCACCGGTGGAGA 185467

QY 602 CCCGATACGAACCTCAACGTGCGCGCGGTGACGTCGTGCGGACCGCGGAAGATACGCG 661

Db 185468 TCACCGAGCATGTGTGAGAGACAAAGGTGTCGACCCGCGCGGTGTTCGAGAGCTCTTG 185527

QY 662 GGGGCTGCTGCCCGCGAGGGGACCAAGTTCTGCGTGACACAGACTCCGTCTTGAAAG 721

Db 185528 CCGGGTTCGAGGACGCG-----TTCTGCTCGATG 185557

QY 722 GGGCCTCGCGGCTTCTCTCTCTCGGAGAGACCGGGGCGCGCGCGCGGATACCTACCT 781

Db 185558 ATCCGACGGGACACACTATATGGGTGATGGCTCCGGGCGGACATGACGATTCGACCC 185617

QY 782 ACCGCGTCGCGACGCGGTGTCCTCCGTCGCGGCTTCGACGGCACACGACCCCGAGCG 841

Db 185618 ACCGGGTGGGGAGGGGAGCTGTTGCACTG----- 185648

QY 842 GCGGCCCCCTTTCAACTACCTGAGGAGACGACTCGAACGCGGACGCGGTCCTCCGTCGCC 901

Db 185649 -----GCTCGTATGATCTGCGTCCCAACCGCGTGGCCC 185683

QY 902 CCGAATGCGCTTCGAGTTCAACTCGGCTACGTCGGGCTTAACTCGGCTACGAGCTGTAAG 961

Db 185684 CGGGGATG-----GATTTCGCGCTGGGGTGGGTGGGATATCTCGGGATGAGATATAAG 185737

QY 962 CGGAGACCAACCGGCGACCCCGCGCACCGGTTCGCCGACCCCGACGCGCGATTCTCTTG 1021

Db 185738 CCGAATGCGGGGTGAGCAATTCGGACGCTCATTCGATCCGATGCCCCATGATCTTTGG 185797

QY 1022 CCGACCGCGCCATCGCCCTCGACACACAGAAAGCTGCTACTCMGTGGCCCTTCGACC 1081

Db 185798 CCGACCGGGCCATTCGCAATCGAAACC-----GCGCGGTGTGGCTCATGCGGCTCGGTG 185851

QY 1082 GCGGGGGCCACGACGAGCGGCGCGCGCTGTGGGTGGGAGAGCGCGGACGACCTTACCG 1141

Db 185852 AGCAGAGGGAG-----TGTGTCCGAGATGATCCGCGGCTCT----- 185888

QY 1142 GCCTGACCGTTCGCGCGCCCGCGGACGAGCCGACCCCGCCATNGATCTTCGGGATCCCGAGG 1201

Db 185889 ---GGGCGAGCTGCGGCCACCCGTTGC----- 185913

QY 1202 CGGCGCGCGGCTTTCGCGCCCTTCGCGCGCGCGGCGGACACAGACGCGCTACTTCAAGC 1261

Db	185914	--GTTGCGCGCCCCGGCGCCGACGTTCAACCTCCGGGATGACCGGACAGCTTACCTGGACA	185971
Qy	1262	GCATGACGAGTGGCTTCAAGAGATCCGCACAGGGGAGTCGTACGATCTGCTTGACCA	1321
Db	185972	TGATGGCCCGGCCCAAGAGATTGATTCACCCGGGGGAGTCTATAGATCTGTGCTGACCA	186031
Qy	1322	ACATGGTCAACGGCGCCGACCGAGGGCGACGGCCCTGCGCTTACTTCGCGCTGGCGCCA	1381
Db	186032	CCACACTGGCGCGGAGGTGAGGTGATCCCTCTGCGCCGCTTATCTHGGGCTGCGGGCGG	186091
Qy	1382	TCAGCGCCGTCGCCGTAACGGCGCCCTGCTCGATTTCGCCAATCTGTGGGTGAGGGCGCT	1441
Db	186092	CCAACCCCACTCTTATGATGATTCCTGCACTGGGGGAGATGGCGGTGTGAGCTTGGT	186151
Qy	1442	CGCCCGAGCGGTTTCTCTCAAGATCGCGCGCCGACGCGGCGCTGCAGTCCAAAGCCATCAAG	1501
Db	186152	CGCGGGAAGGTTTCAACATCGAAGCTTCGGAAGCTTCGGAATCCAAACCCATCAAG	186211
Qy	1502	GGACCCGCCCCCGGGGCGGACCGCGGAGAGAGACGAGCGGCTCCGCGCTGACCTGGCCG	1561
Db	186212	GGACCCGCCCCCGGGGCGAGTACCGAGACGAGAGATGGCGGCTTGTATCGCGGATCTTCACG	186271
Qy	1562	GGCGGGAGAGAGACCGGGCGGAGAACTGTATGTGTGACCTGTGGTCCGCAACGACTCA	1621
Db	186272	AACAACCCCAAGGACCGCGCGGAACTCTATATGTGTGACCTGTGGTCCGCAATGATCTTG	186331
Qy	1622	ACAGCGTCTGCGCGATTCGGCTCCGTCCACGTCGCCCGGCTCTTGAGGTGAGAACCTAG	1681
Db	186332	CCCCGGGGGCGCAACCGGCAACGTGTGAGTGGAGAAAGCTTTTGACGTGAGACCTTAG	186391
Qy	1682	CGCCGTGACCAAGCTGTGTGACCAATCCGGGGAGCGGCTGGCGCCCGGACACGACCG	1741
Db	186392	CCACCGTCCACAGCTGTGTGACCAACATTAACGCCACGCTGGAG--GCTAAGAACCCCA	186448
Qy	1742	CCGCTGTGTACGGCGCCGCTTCCCGGGGCTGCATGACCGGGCGGCCCAAGACCGCA	1801
Db	186449	TTGACTGTGTGTGGGGCGGGGTTCCCGGTGTTCATAGCCGGTGCACCGAAGATCCGA	186508
Qy	1802	CCATGAGATCATTCGACCGCTGAGAGAAAGCCCCCGGGGCGCTTACTCCGGGCGCTCG	1861
Db	186509	CCATGGAATCATTCGACGAGCTGAGAGACCGGCCCTCGTGTATCTATTCCGGTGGCTGG	186588
Qy	1862	GATGTTGCGCCTCTACGGCGCGCGCGCACTTCAGATGTGTATCCGACCATGTGCTGG	1921
Db	186589	GCTATTTCTCCCTTATGATGGGGGGGTGATCTGTGATGATGATCCGACCGTGTGTCTA	186628
Qy	1922	CCGACGGCCACGGGAGTTTGGCGCTGGCGGGGCGATGTGTCTCTTCCGACGAGAGG	1981
Db	186629	CCCCCGGCTCTGTGAATGAGGGGTGGCGGGGCGATCTTGCGCTGTGAGACCGCGCG	186688
Qy	1982	AGAGTTTCAACGAGACGTTGTAAAGGCCCCGGCGCATGGT 2021	
Db	186689	CAGAGTGGAGAGATCCGGGTGAATGAGGGCCCTGCT 186728	
RESULT 15			
A48326	LOCUS	A48326	4496 bp
DEFINITION	Sequence 4 from Patent WO9601901.	DNA	linear
ACCESSION	A48326		PAT 07-MAR-1997
VERSION	A48326.1	GI:2302119	
KEYWORDS			
SOURCE			
ORGANISM	Streptomyces pristinaespiralis		
REFERENCE	Streptomyces pristinaespiralis		
AUTHORS	Streptomyces pristinaespiralis		
	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
	Streptomycinae; Streptomycetaceae; Streptomyces.		
	1 (bases 1 to 4496)		
	Blanc, Y., Thibaut, D., Bamae-Jacques, N., Blanche, F., Crouzet, J.,		
	Barriere, J., Debussche, L., Fanechon, A., Paris, J. and		
	Duclercq-Rosset, G.		
	STREPTOGRAMINS AND METHOD FOR PREPARING SAME BY MUTASYNTHESIS		
	Patent: WO 9601901-A 4 25-JAN-1996;		
	PHONE POULENC RORER SA (FR)		







This Page Blank (uspto)







CC The present invention relates to a transformant having been modified so  
CC as to produce a secondary metabolite. The secondary metabolite has a  
CC benzene ring skeleton free from substitution at the para-position by a  
CC nitrogen-containing functional group, thereby enabling the production of  
CC a secondary metabolite with a benzene ring skeleton substituted at the  
CC para-position by a nitrogen-containing group. The transformant organism  
CC participating in the biosynthesis pathway from chorismic acid into p-  
CC anthonylpyruvic acid. The present sequence is the coding sequence for  
CC 4-amino-4-deoxychorismic acid synthase (papa), from *Streptomyces*  
CC *venezuelae*. papa participates in the biosynthesis pathway from chorismic  
CC acid into p-anthonylpyruvic acid, and so the papa gene can be used to  
CC produce the transformant of the present invention. The transformant can  
CC be used to produce metabolites for application in pharmaceuticals,  
CC veterinary drugs and agrochemicals

XX  
SO Sequence 2061 BP; 297 A; 842 C; 654 G; 268 T; 0 U; 0 Other;

Query Match 100.0%; Score 2061; DB 4; Length 2061;  
Best Local Similarity 100.0%; Pred. No. 1.7e-253;  
Matches 2061; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGACGCTTCGATCGACAACTAGACTGTTACCCACAACCTGTTCCAGTATATC 60  
DB 1 ATGCGACGCTTCGATCGACAACTAGACTGTTACCCACAACCTGTTCCAGTATATC 60  
QY 61 GCGGAGGCCACCGGAGCAACCCCGCTGCTGTCGCCAACGACGCCGCACTGTCGCGCTG 120  
DB 61 GCGGAGGCCACCGGAGCAACCCCGCTGCTGTCGCCAACGACGCCGCACTGTCGCGCTG 120  
QY 121 CCGGTGAGGACTTTCGACGCGGATGTCGTGTCCTCCGCGGCCCGGACGCCGAGCCGAGAA 180  
DB 121 CCGGTGAGGACTTTCGACGCGGATGTCGTGTCCTCCGCGGCCCGGACGCCGAGCCGAGAA 180  
QY 181 CCGGACTTCGGAATTCAGCCGCCCGGCGGATATACGACAGCGGCTGCTGTCGCGCTC 240  
DB 181 CCGGACTTCGGAATTCAGCCGCCCGGCGGATATACGACAGCGGCTGCTGTCGCGCTC 240  
QY 241 TGCTCGGACCAACGAGGATGCGCCAGCTCTTCGCGCGGAAACGTCGCTGCTGCGCGGAA 300  
DB 241 TGCTCGGACCAACGAGGATGCGCCAGCTCTTCGCGCGGAAACGTCGCTGCTGCGCGGAA 300  
QY 301 CCGATGACGCGCGGCTTCGAGGTGCGGACACCGGCGAGGACGTCCTCCGCGGCGCTC 360  
DB 301 CCGATGACGCGCGGCTTCGAGGTGCGGACACCGGCGAGGACGTCCTCCGCGGCGCTC 360  
QY 361 CCGCTGCGCTTCACCGCGCTGCGCTACACTCCCTGCGGCCGACCGACTCCCGGAGAG 420  
DB 361 CCGCTGCGCTTCACCGCGCTGCGCTACACTCCCTGCGGCCGACCGACTCCCGGAGAG 420  
QY 421 CTGGAACCCCTCGCTGAGAGCGACGAGGAGTGTATGAGGCTGCGGCGACCGCGAGAG 480  
DB 421 CTGGAACCCCTCGCTGAGAGCGACGAGGAGTGTATGAGGCTGCGGCGACCGCGAGAG 480  
QY 481 CCGCTGAGGCGCTTCAGTTCCACCGGAGTTCATCGGACGCACTTCGCGCGGAGATC 540  
DB 481 CCGCTGAGGCGCTTCAGTTCCACCGGAGTTCATCGGAGCGCACTTCGCGCGGAGATC 540  
QY 541 ATGCGCAACTTCGCGGACCTGCGCTTGCCCAACCGGCGACGCGGCGCCGCGGCGCGAC 600  
DB 541 ATGCGCAACTTCGCGGACCTGCGCTTGCCCAACCGGCGACGCGGCGCCGCGGCGCGAC 600  
QY 601 TCCCGTGAAGAACTCCACGTCGCGCGGTGAGCTGTCGCGGACCGCGAAGAGTACGC 660  
DB 601 TCCCGTGAAGAACTCCACGTCGCGCGGTGAGCTGTCGCGGACCGCGAAGAGTACGC 660  
QY 661 CCGGCTGCTGCTGCGCGGACGACCACTTCGCTGAGACAGACTTCGCTCTCGAA 720  
DB 661 CCGGCTGCTGCTGCGCGGACGACCACTTCGCTGAGACAGACTTCGCTCTCGAA 720  
QY 721 GCGGCTGCTGCTGCTTCCTTCCTGCGGACGACCGCGGCGCGCTGCGGAGTACCTTACC 780  
DB 721 GCGGCTGCTGCTGCTTCCTTCCTGCGGACGACCGCGGCGCGCTGCGGAGTACCTTACC 780

QY 781 TACCGGTGCGCGACGCGGCTGCTCCGTCGCGGCTCCGACGCGACCAACCGCGAGC 840  
DB 781 TACCGGTGCGCGACGCGGCTGCTCCGTCGCGGCTCCGACGCGACCAACCGCGAGC 840  
QY 841 CCGCGCTCTTCTTCACTGAGAGACAGCTGAAACCGCGAGGCTCCGCTGCC 900  
DB 841 CCGCGCTCTTCTTCACTGAGAGAGAGCTGAAACCGCGAGGCTCCGCTGCC 900  
QY 901 CCGGAATGCGCTTCGAGTTCAACCTCGGCTACGTGGCTTACCTGGGCTTACGAGCTGAAG 960  
DB 901 CCGGAATGCGCTTCGAGTTCAACCTCGGCTACGTGGCTTACCTGGGCTTACGAGCTGAAG 960  
QY 961 GCGGAGACCAACCGGACCGCGGACCGGCTCCCGACCGCGACCGCGGCTCTTC 1020  
DB 961 GCGGAGACCAACCGGACCGCGGACCGGCTCCCGACCGCGACCGCGGCTCTTC 1020  
QY 1021 GCGGACCGGCAATGCGCTTCGACACGAGAAAGCTGCTACTCTGTGCGCTTCGAC 1080  
DB 1021 GCGGACCGGCAATGCGCTTCGACACGAGAAAGCTGCTACTCTGTGCGCTTCGAC 1080  
QY 1081 CCGCGGCGCACGACGACGCGCGCTGCGCTGCGGAGACGCGGACCGCTTACC 1140  
DB 1081 CCGCGGCGCACGACGACGCGCGCTGCGCTGCGGAGACGCGGACCGCTTACC 1140  
QY 1141 GCGCTGCGCTGCGCGCTCCCGGCGGACCGCGGACCGCGGCTTCCGAGATCCCGAG 1200  
DB 1141 GCGCTGCGCTGCGCGCTCCCGGCGGACCGCGGACCGCGGCTTCCGAGATCCCGAG 1200  
QY 1201 GCGGCGGCGGCTTCGCGCTTCGCGCTGCGCGGCGGACGACGACGAGCGCTTCAAG 1260  
DB 1201 GCGGCGGCGGCTTCGCGCTTCGCGCTGCGCGGCGGACGACGAGCGCTTCAAG 1260  
QY 1261 CCGATGAGAGAGTCTTCAAGAGATTCGCGAAACGAGAGTGTGATGATTCGCTGACC 1320  
DB 1261 CCGATGAGAGAGTCTTCAAGAGATTCGCGAAACGAGAGTGTGATGATTCGCTGACC 1320  
QY 1321 AACATGATCAACGCGCGGACCGAGGCGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380  
DB 1321 AACATGATCAACGCGCGGACCGAGGCGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380  
QY 1381 ATGAGCCCGCTCCGTAACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440  
DB 1381 ATGAGCCCGCTCCGTAACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440  
QY 1441 TCGCCGGAACGCTTCTTCAAGATCGGCGCGGACGCGGCGGCTGAGTCCAAAGCCATCAAG 1500  
DB 1441 TCGCCGGAACGCTTCTTCAAGATCGGCGCGGACGCGGCGGCTGAGTCCAAAGCCATCAAG 1500  
QY 1501 GCGACCCGCGCGCGGCGGCGGACCGGCGGAGAGAGAGAGCGGCTGCGGCGGACCGGCTG 1560  
DB 1501 GCGACCCGCGCGCGGCGGCGGACCGGCGGAGAGAGAGAGCGGCTGCGGCGGACCGGCTG 1560  
QY 1561 GCGCGGAGAGAGAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620  
DB 1561 GCGCGGAGAGAGAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620  
QY 1621 AACAGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 1680  
DB 1621 AACAGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 1680  
QY 1681 GCGCGCTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740  
DB 1681 GCGCGCTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740  
QY 1741 GCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800  
DB 1741 GCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800  
QY 1801 ACCATGAGATATGACGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860  
DB 1801 ACCATGAGATATGACGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860







QY	1081	GGCCGGGGGCAACGAGACGGCGGCCCGGCGCTGTGCTGGGAGAGAGGCCGAGACCTTCAAC	1140
Db	1081	CGCCGGGGGCAACGAGAGGGCGGCCCGCGCCTGTGCTGGGAGAGAGGCCGAGACCTTCAAC	1140
QY	1141	GGCCGTGGCGGTTCGGCGGCCCGGAGCGAGCGACCCCGGCATAGGTCTTTGGGATCCCGAG	1200
Db	1141	GGCCGTGGCGGTTCGGCGGCCCGGAGCGAGCGACCCCGGCATAGGTCTTTGGGATCCCGAG	1200
QY	1201	GGGGCGGCGGCTTCGGGCTCCCTTGCGCCGCGCGCGCACGACAAGAGCGCTTACCTCAAG	1260
Db	1201	GGGGCGGCGGCGCTTCGGGCTCCCTTGCGCCCGCGCGCGCACGACAAGAGCGCTTACCTCAAG	1260
QY	1261	CGCATCGACGAGTGTCTTAAGAGATTCGGCAACGGCGAGTGTGTACGAGATTTGGCTTACAC	1320
Db	1261	CGCATCGACGAGTGTCTTAAGAGATTCGGCAACGGCGAGTGTGTACGAGATTTGGCTTACAC	1320
QY	1381	AACATGTGTACCGCGCGCGACGAGGCGAGCGGCCCTCGCGCTCTTACTCGCGCTCGCGCC	1440
Db	1381	ATCAGCCCCGCTCCGCTACGCGCGCCCTGCTCGAGTTCCCGAATCTGTGGTCTGAGCGCC	1440
QY	1441	TGCGCCGAGCGGTTCTCTACGATCGGGGCGCACCGCGCGCGTGTGATCCAAACCATCAAG	1500
Db	1441	TGCGCCGAGCGGTTCTCTACGATCGGGGCGCACCGCGCGCGTGTGATCCAAACCATCAAG	1500
QY	1501	GGGACCGCGCCCCCGGGGGCGGACCGCGGAGGAGAGCGACCGGCTCCGCGCCGACCTTGACC	1560
Db	1501	GGGACCGCGCCCCCGGGGGCGGACCGCGGAGGAGAGCGACCGGCTCCGCGCCGACCTTGACC	1560
QY	1561	GGCCGGGAGAAAGAACCGGGCCGGAACCTGTATGTGTGACCTGTGTCCGCAAGACTTC	1620
Db	1561	GGCCCGGAGAAAGAACCGGGCCGGAACCTGTATGTGTGACCTGTGTCCGCAAGACTTC	1620
QY	1621	AACAGCGATCGGCGCATCGGCTCCGCTCCAGCGTCCCGGCTCTTCGAGGTGGAAGCCTAC	1680
Db	1621	AACAGCGATCGGCGCATCGGCTCCGCTCCAGCGTCCCGGCTCTTCGAGGTGGAAGACTTAC	1680
QY	1681	GGCGCCGTGCACACAGCTGTGTGCACCATTCGGGGACGGCTGCGGCGCCGAGCACGACACC	1740
Db	1681	GGCGCCGTGCACACAGCTGTGTGCACCATTCGGGGACGGCTGCGGCGCCGAGCACGACACC	1740
QY	1741	GGCGCTTCGGTACCGGCGCGCTTCCCGGGCGGCTTCATGACCGGCGCGGCCCAAGAGCGC	1800
Db	1741	GGCGCTTCGGTACCGGCGCGCTTCCCGGGCGGCTTCATGACCGGCGCGGCCCAAGAGAGCGC	1800
QY	1801	ACCATGAGATCATGAGACCGGCTGGAGGAAGCCCCCGGGGCGTCTACTCCGGGGCGGCTC	1860
Db	1801	ACCATGAGATCATGAGACCGGCTGGAGGAAGCCCCCGGGGCGTCTACTCCGGGGCGGCTC	1860
QY	1861	GGATGTTCGCGCTCAGCGGCGCGCGCGACCTCAGCATCGTCAATCCGACCATGTGTGTG	1920
Db	1861	GGATGTTCGCGCTCAGCGGCGCGCGCGACCTCAGCATCGTCAATCCGACCATGTGTGTG	1920
QY	1921	GGCGACGGGCGAGCGGAGTTTGGCGGTGGGGGGGAGATGTGTCTTCCGACCAAGAG	1980
Db	1921	GGCGACGGGCGAGCGGAGTTTGGCGGTGGGGGGGAGATGTGTCTTCCGACCAAGAG	1980
QY	1981	GAGAGATTCAACCGAGACCGTGTAAAGCGCGCGCATGTGTACCGGCTTCGACGGGAGC	2040
Db	1981	GAGAGATTCAACCGAGACCGTGTAAAGCGCGCGCATGTGTACCGGCTTCGACGGGAGC	2040
QY	2041	GGCGTGGCGGGGCGCCGATGA 2061	
Db	2041	GGCGTGGCGGGGCGCCGATGA 2061	

RESULT	3
ABZ69799	
ID	ABZ69799 standard; DNA; 3305 BP.

XX	AB269799;
AC	
XX	
DT	08-APR-2003 (first entry)
XX	
DE	Plasmid papABC.
XX	
KM	DHFR; translation; orthogonal tRNA; O-tRNA; O-RS; CAT; TyrRS;
KW	orthogonal aminoacyl tRNA synthetase; unnatural amino acid;
KX	chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;
KM	biosynthesis; p-aminophenylalanine; pAF; chorismate.
XX	
OS	Synthetic.
XX	
PN	WO200205923-A2.
XX	
PD	31-OCT-2002.
XX	
PF	19-APR-2002; 2002WO-US012465.
XX	
PR	19-APR-2001; 2001US-0285030P.
PR	06-FEB-2002; 2002US-0355514P.
XX	
PA	(SCRI ) SRIIPS RES INST.
XX	
PI	Schultz P, Wang L, Anderson JC, Chin JWK, Liu DR, Magliery TJ;
PI	Meggere EL, Mehl RA, Pasternak M, Santoro SW, Zhang Z;
XX	
DR	WPI, 2003-120430/11.
PT	
PT	Composition useful for producing protein comprising unnatural amino acid,
PT	has translation system comprising orthogonal tRNA and orthogonal
PT	aminoacyl tRNA synthetase.
XX	
PS	Example 4; Page 127-128; 188pp; English.
XX	
CC	The invention relates to a novel composition comprising a translation
CC	system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl
CC	tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O
CC	-tRNA with at least one unnatural amino acid in the translation system
CC	and the O-tRNA recognises at least one selector codon. A composition of
CC	the invention is useful for producing at least one protein comprising at
CC	least one unnatural amino acid. The protein is the Asp17AG mutant of
CC	chloramphenicol acetyltransferase (CAT), the Tyr163TAG mutant of mouse
CC	dihydrofolate reductase (DHFR), or a Tyr163TAG mutant of mouse
CC	dihydrofolate reductase comprising CHOCH His6tag. The unnatural amino
CC	acid is provided exogenously. The translation system is a cell and the
CC	unnatural amino acid is biosynthesised by the cell. The present sequence
CC	represents a plasmid containing the individual genes papABC that encode
CC	the enzymes used to carry out the conversion of chorismate to the
CC	unnatural amino acid p-aminophenylalanine (pAF)
XX	
XX	
SO	Sequence 3305 BP; 445 A; 1327 C; 1121 G; 412 T; 0 U; 0 Other;
	Query Match
	Best Local Similarity 92.8%; Score 1773.8; DB 8; Length 3305;
	Matches 1913; Conservative 0; Mismatches 97; Indels 51; Gaps 3;
OY	1 ATGGCGCAGCTTTCGATCGACAATTAGACTCCGTTCACCACCAACTCTTTCCAATACATC 60
DB	1 ATGGCGCAGCTTTCGATCGACAATTAGACTCCGTTCACCACCAACTCTTTCCAATACATC 60
OY	GAGCGAGGCACCGGAGCAACCCCCCGTGCGTGTGCCAAGCAGCGCGCATGGTGGGGCTG 120
DB	61 GAGCGAGGCACCGGAGCACCCCCCG---TCGTGCCCAAGCAGCGCGCATGGTCCGGCTG 117
OY	CCCCTCGAGGACTTCGACGCGCATGTCGTGTCCCGGAGCCCGGACAGCCCGACCGGAAA 180
DB	118 CCCCTCGAGGACTTCGACGCGCATGTCGTGTGTCCCGGAGCCCGGACAGCCCGGAAA 177
OY	CGGAGACTTCGGAAATCAGCCGCGCGGCGATACCGACAGCGGCGTCCCGTCTTGCGCGTC 240
DB	181 CGGAGACTTCGGAAATCAGCCGCGCGGCGATACCGACAGCGGCGTCCCGTCTTGCGCGTC 240
OY	CGGAGACTTCGGAAATCAGCCGCGCGGCGATACCGACAGCGGCGTCCCGTCTTGCGCGTC 237
DB	178 CGGAGACTTCGGAAATCAGCCGCGCGGCGATACCGACAGCGGCGTCCCGTCTTGCGCGTC 237



```

QY 241 TGCTCGGACCAACGAGGATCGCCAGCTTTGCGGAGAACCGTGGGCTGCCCCGAAA 300
Db 238 TGCTCGGACCAACGAGGATCGCCAGCTT-----CGGAGAA 276
QY 301 CCATGACGAGCGCGGATCTCCAGATGCGGACACCGGAGAGAGCTTTCCGAGGCTC 360
Db 277 CCATGACGAGCGCGGATCTCCAGATGCGGACACCGGAGAGAGCTTTCCGAGGCTC 336
QY 361 CCTTCGCGCTTCAACCGCGCTGCGTACCTCCCTGGCGGACCGACTTCCCGAGAG 420
Db 337 CCTTCGCGCTTCAACCGCGCTGCGTACCTCCCTGGCGGACCGACTTCCCGAGAG 396
QY 421 CTGAAACCCCTCGCTGAGAGGACGAGCGGGGTGTCATGGGGCTGGGAGACCGGAGAG 480
Db 397 CTGAAACCCCTCGCTGAGAGGACGAGCGGGGTGTCATGGGGCTGGGAGACCGGAGAG 456
QY 481 CCGCTGAGGAGCGCTCAAGTTCACCCGAGATCATCGGACGACTTTCGGCGGAGATC 540
Db 457 CCGCTGAGGAGCGCTCAAGTTCACCCGAGATCATCGGACGACTTTCGGCGGAGATC 516
QY 541 ATGAGCAACTTTCGCGACCTCGCTTCCGACCAACCGGAGACCGGCGCACGGGGCCGAC 600
Db 517 ATGAGCAACTTTCGCGACCTCGCTTCCGACCAACCGGAGACCGGCGCACGGGGCCGAC 576
QY 601 TCCCGTACGAACTCAAGTGGCGCGGAGTGCATGGCGGAGACCGGAGAGGTACGC 660
Db 577 TGGAGCTACGAACTCAAGTGGCGCGGAGTGCATGGCGGAGACCGGAGAGGTACGC 636
QY 661 CCGGCTGCTGCGCGGAGGAGGACCACTTTCGCTGACACAGCACTCCGCTCTCGAA 720
Db 637 CCGGCTGCTGCGCGGAGGAGGACCACTTTCGCTGACACAGCACTCCGCTCTCGAA 696
QY 721 GGGGCTTCGCGCTTCTCTTCTCGGAGACGCGCGGCGGCTCGCGAGTACTAC 780
Db 697 GGGGCTTCGCGCTTCTCTTCTCGGAGACGCGCGGCGGCTCGCGAGTACTAC 756
QY 781 TACCGGTGCGGAGGAGGAGTCTCGTCCGCGGCTCGACCGGACACGACCGGAGC 840
Db 757 TACCGGTGCGGAGGAGGAGTCTCGTCCGCGGCTCGACCGGACACGACCGGAGC 816
QY 841 CCGGCGCTTCTTCACTACCTGAGAGAGTCTGACGCGGAGCGGATCCCGTTCGC 900
Db 817 CCGGCGCTTCTTCACTACCTGAGAGAGTCTGACGCGGAGCGGATCCCGTTCGC 876
QY 901 CCGGAATGCGCTTTCAGATTCAACTTGGCTACGCTGAGTACCTCGGCTACGAGTGA 960
Db 877 CCGGAATGCGCTTTCAGATTCAACTTGGCTACGCTGAGTACCTCGGCTACGAGTGA 936
QY 961 GGGGAGACACCGGCGGAGCGGCGGACCGGATCCCGGACCGGCGGCTTCTTC 1020
Db 937 GGGGAGACACCGGCGGAGCGGCGGACCGGATCCCGGACCGGCGGCTTCTTC 996
QY 1021 GCGGACCGGCGCATCGGCTTTCAGACCAACGAGAGGCTGCTACCTGCTGAGCTTCGAC 1080
Db 997 GCGGACCGGCGCATCGGCTTTCAGACCAACGAGAGGCTGCTGCTGAGCTTCGAC 1056
QY 1081 CCGCGGAGGACGACGAGCGGCGGCGGCTGCTGAGGAGACCGGCGGACCTTAC 1140
Db 1057 CCGCGGAGGACGACGAGCGGCGGCGGCTGCTGAGGAGACCGGCGGACCTTAC 1116
QY 1141 GGGCTGCGGCTCGCGCGCGGCGGAGCGGACCGGCTGCTTTCGAGATCCCGAG 1200
Db 1117 GGGCTGCGGCTCGCGCGCGGCGGAGCGGACCGGCTGCTTTCGAGATCCCGAG 1176
QY 1201 GCGGCGGCGGCTTTCGCGCGGCTGAGCGGCGGCGGACGACGAGAGGCTGCTCAAG 1260
Db 1177 GCGGCGGCGGCTTTCGCGCGGCTGAGCGGCGGCGGACGACGAGAGGCTGCTCAAG 1228
QY 1261 CGCATGACGAGTCTCAAGAGATCGCAACGGGAGTCTGACGATCTGCTGACC 1320
Db 1229 -----CGGCGCTCGCAACGGGAGTCTGACGATCTGCTGACC 1269

```

```

QY 1321 AACATGATCAACGCGGACGACCGAGGAGCGGCGGCTTCTACTCGGCGTGGCGCC 1380
Db 1270 AACATGATCAACGCGGACGACCGAGGAGCGGCGGCTTCTACTCGGCGTGGCGCC 1329
QY 1381 ATGAGCCCGCTCCGATACGAGCGGCGGCTGCTGAGTTCGCCAATGCTGAGTGGCGCC 1440
Db 1330 ATGAGCCCGCTCCGATGAGCGGCGGCTGCTGAGTTCGCCAATGCTGAGTGGCGCC 1389
QY 1441 TCGCCGAGGCGGCTTCTCAAGATCGGCGGCGGAGCGGCGGCTGCTGAGTTCGAGTCAAG 1500
Db 1390 TCGCCGAGGCGGCTTCTCAAGATCGGCGGCGGAGCGGCGGCTGCTGAGTTCGAGTCAAG 1449
QY 1501 GGGACCGGCGGCGGCGGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
Db 1450 GGGACCGGCGGCGGCGGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1509
QY 1561 GGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
Db 1510 GGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1569
QY 1621 AACAGGCTTTCGCGGATCGGCTTCCGACGAGTGCCTCGGCTTTCGAGTGGAGACTTAC 1680
Db 1570 AACAGGCTTTCGCGGATCGGCTTCCGACGAGTGCCTCGGCTTTCGAGTGGAGACTTAC 1629
QY 1681 GCGCGGCTGACGAGCTGCTGACCATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
Db 1630 GCGCGGCTGACGAGCTGCTGACCATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1689
QY 1741 GCGGCTGCTGACGAGCTTCCCGGCGGCTTCCGACGAGTGCCTCGGCTTTCGAGTGGAGACTTAC 1800
Db 1690 GCGGCTGCTGACGAGCTTCCCGGCGGCTTCCGACGAGTGCCTCGGCTTTCGAGTGGAGACTTAC 1749
QY 1801 ACCATGAGATCATCAACGCGCTGAGAGAGGCGGCGGCGGCTTTCGAGTGGAGACTTAC 1860
Db 1750 ACCATGAGATCATCAACGCGCTGAGAGAGGCGGCGGCGGCTTTCGAGTGGAGACTTAC 1809
QY 1861 GGAATGTTGCGGCTTTCAGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920
Db 1810 GGAATGTTGCGGCTTTCAGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1869
QY 1921 GCGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980
Db 1870 GCGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1929
QY 1981 GAGGAGTTCAACGAGACCGTGTAAAGGCGGCGGAGTGTACCGGCTTTCGAGTGGAGACTTAC 2040
Db 1930 GAGGAGTTCAACGAGACCGTGTAAAGGCGGCGGAGTGTACCGGCTTTCGAGTGGAGACTTAC 1989
QY 2041 GCGGAGGCGGCGGCGGAGTGA 2061
Db 1990 GCAGTGCGGCGGCGGAGTGA 2010

```

RESULT 4  
 ABZ69798  
 ID ABZ69798 standard; DNA; 12391 BP.  
 XX ABZ69798;  
 AC  
 XX  
 DT 08-APR-2003 (first entry)  
 XX  
 DE Plasmid pasc-papabc.  
 XX  
 XX DHFR; translation; orthogonal tRNA; O-tRNA; O-RS; CAT; TyRS;  
 KM orthogonal aminoacyl tRNA synthetase; unnatural amino acid;  
 KM chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;  
 KM biosynthesis; p-antimphenylalanine; paf.  
 OS Synthetic.  
 XX  
 PN MO200285923-A2.  
 XX  
 PD 31-OCT-2002.



XX 19-APR-2002; 2002M0-US012465.  
XX  
PF 19-APR-2001; 2001US-0285030P.  
PR 06-FEB-2002; 2002US-0355514P.  
XX  
XX (SCRI ) SCRIPPS RES INST.  
XX  
PI Schultz P, Wang L, Anderson JC, Chin JWK, Liu DR, Magliery TJ,  
PI Meggers EL, Mehl RA, Pasternak M, Santoro SW, Zhang Z,  
XX  
XX WPI; 2003-120430/11.  
XX  
XX  
XX Composition useful for producing protein comprising unnatural amino acid,  
XX has translation system comprising orthogonal tRNA and orthogonal  
XX aminoacyl tRNA synthetase.  
XX  
XX Example 4; Page 124-127; 188pp; English.  
XX  
XX The invention relates to a novel composition comprising a translation  
XX system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl  
XX tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O  
XX -tRNA with at least one unnatural amino acid in the translation system  
XX and the O-tRNA recognises at least one selector codon. A composition of  
XX the invention is useful for producing at least one protein comprising at  
XX least one unnatural amino acid. The protein is the Asp12TAG mutant of  
XX chloramphenicol acetyltransferase (CAT), the Tyr163TAG mutant of mouse  
XX dihydrofolate reductase (DHFR), or a Tyr163TAG mutant of mouse  
XX acid is provided exogenously. The translation system is a cell and the  
XX unnatural amino acid is biosynthesised by the cell. The present sequence  
XX represents a plasmid for use in the biosynthesis of p-aminophenylalanine  
XX (PAF) in vivo  
XX  
SQ Sequence 12391 BP; 2830 A; 3588 C; 3346 G; 2627 T; 0 U; 0 Other;  
Query Match 86.0%; Score 1772.2; DB 8; Length 12391;  
Best Local Similarity 92.8%; Pred. No. 5.5e-217;  
Matches 1912; Conservative 0; Mismatches 98; Indels 51; Gaps 3;  
QY 1 ATGCGACGCTTCTGATCGACAACTAGTACTGTTCCACCAACCTGTTCCAGTACATC 60  
DB 24 ATGCGACGCTTCTGATCGACAACTAGTACTGTTCCACCAACCTGTTCCAGTACATC 83  
QY 61 GGGAGGCGCACCGGCGCAACCCCGCTGCTGTCGCCAAGCGCGCACTGTCGGCGCTG 120  
DB 84 GGGAGGCGCACCGGCGCAACCCCGCTGCTGTCGCCAAGCGCGCACTGTCGGCGCTG 140  
QY 121 CCGCTCGAGGACTTTCGACGCGATGCTGTCGCCCGGCGCGCGACCGCGCAACCGGAA 180  
DB 141 CCGCTCGAGGACTTTCGACGCGATGCTGTCGCCCGGCGCGCGACCGCGCAACCGGAA 200  
QY 181 CCGGACCTTCGGAATCAGCGCGCGCGCGCGATCAGCGACGCGCGCTGCCCTTCCTCGGCTC 240  
DB 201 CCGGACCTTCGGAATCAGCGCGCGCGCGCGATCAGCGACGCGCGCTGCCCTTCCTCGGCTC 260  
QY 241 TGCTCGGCGCACGAGGCGCATCGCCAGCTCTTGCGCGGAACCGTCGGCTCGCCCGGAA 300  
DB 261 TGCTCGGCGCACGAGGCGCATCGCCAGCTCTTGCGCGGAACCGTCGGCTCGCCCGGAA 320  
QY 301 CCGATGACGCGCGCGGCTCTCGAGGTGCGGACACCGCGCGAGGACGTCCTTCGCGGCGCTC 360  
DB 300 CCGATGACGCGCGCGGCTCTCGAGGTGCGGACACCGCGCGAGGACGTCCTTCGCGGCGCTC 359  
QY 361 CCGCTCGCGTTTCAACCGCGCGTTCACCTCCCTGGCGCGCACCGACCTTCCTCCGAGAG 420  
DB 360 CCGCTCGCGTTTCAACCGCGCGTTCACCTCCCTGGCGCGCACCGACCTTCCTCCGAGAG 419  
QY 421 CTGGAACCTCTCGCGTTCGAGGACGACGAGGGGTGTCTATGGGCTTCGCGGCAACCGCGAAG 480  
DB 420 CTGGAACCTCTCGCGTTCGAGGACGACGAGGGGTGTCTATGGGCTTCGCGGCAACCGCGAAG 479  
QY 481 CCGCTGTGGGCGCTTCAGTTCCACCGGAGTCCATCGGACGACGACTTCGCGCGGAGATC 540

DB 480 CCGCTATGAGCGCTTCAGTTCCACCGGAGTCCATCGGACGCACTTCGCGCGGAGATC 539  
QY 541 ATGCGCACTTCGCGGACCGCGCTCGCGCACCAACCGGCGACGCGCGCACCGGCGCGAC 600  
DB 540 ATGCGCACTTCGCGGACCGCGCTCGCGCACCAACCGGCGACGCGCGCACCGGCGCGAC 599  
QY 601 TCCCGTACGAACCTTCAGGTGCGCGCGGTGACGTCGTGCGGACCGCGGAAGGTAAGC 660  
DB 600 TGGGGCTACGAACCTTCAGGTGCGCGCGGTGACGTCGTGCGGACCGCGGAAGGTAAGC 659  
QY 661 CCGGCTGCTGCGCGCGCGCGGACCAACGTTTGGCTGACGACGCTTCCTTCGAA 720  
DB 660 CCGGCTGCTGCGCGCGCGGACCAACGTTTGGCTGACGACGCTTCCTTCGAA 719  
QY 721 GCGCGTTCGCGCTTCCTTCCTTCGCGGACGACCGCGGCGCGGCTTCGAACTCAGC 780  
DB 720 GCGCGTTCGCGCTTCCTTCCTTCGCGGACGACCGCGGCGCGGCTTCGAACTCAGC 779  
QY 781 TACCGGTGCGCGACCGCGGTGCTTCCTTCGCGGCTTCGACCGGACCAACCGCGAGC 840  
DB 780 TACCGGTGCGCGACCGCGGTGCTTCCTTCGCGGCTTCGACCGGACCAACCGCGAGC 839  
QY 841 CCGCGCGCTTCCTTCACCTGACGAGGACGCTGAAACCGCGAGGTCCTTCGCGC 900  
DB 840 CCGCGCGCTTCCTTCACCTGACGAGGACGCTGAAACCGCGAGGTCCTTCGCGC 899  
QY 901 CCGGACGCGCTTCGAGTTCAACCTCGGCTACGCGGCTTCGAGCTTCGAGCTGAGG 960  
DB 900 CCGGACGCGCTTCGAGTTCAACCTCGGCTACGCGGCTTCGAGCTTCGAGCTGAGG 959  
QY 961 GCGGAGACCAACCGGCGACCGCGCGCGCGCTTCGCGACCGCGCGCTTCCTTC 1020  
DB 960 GCGGAGACCAACCGGCGACCGCGCGCGCGCTTCGCGACCGCGCGCTTCCTTC 1019  
QY 1021 GCGGACCGGCGCATCGCGCTTCGACCAACGAGGCTGCTGCTGCTGCTGCGCTTCGAC 1080  
DB 1020 GCGGACCGGCGCATCGCGCTTCGACCAACGAGGCTGCTGCTGCTGCTGCGCTTCGAC 1079  
QY 1081 CCGCGGCGCGACGAGACGCGCGCGCGCGCTTCGCGGAGGACCGCGGACCGCTCACC 1140  
DB 1080 CCGCGGCGCGACGAGACGCGCGCGCGCGCGCTTCGCGGAGGACCGCGGACCGCTCACC 1139  
QY 1141 GCGCTGCGCGCTTCG 1200  
DB 1140 GCGCTGCGCGCTTCG 1199  
QY 1201 GCGGCGCGCGCTTCG 1260  
DB 1200 GCGGCGCGCGCTTCG 1251  
QY 1261 CCGATGACGAGGCTTCAGGATCGGACGAGGACGAGGATCGGACGAGGATCGGACGAG 1320  
DB 1252 CCGGCTTCGCGGACGAGGATCGGACGAGGATCGGACGAGGATCGGACGAG 1292  
QY 1321 AACATGATCAGCG 1380  
DB 1293 AACATGATCAGCG 1352  
QY 1441 TGCGCGGAGGCTTCCTTCAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500  
DB 1413 TGCGCGGAGGCTTCCTTCAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1472  
QY 1501 GGGACCG 1560  
DB 1473 GGGACCG 1532  
QY 1561 GCGCGGAGGAGGACCGGCGCGGACCTGATGATGTCGACCTGCTTCGCAACGACCTTC 1620



```
Db      1533 GGGCGGAGAGAGACCGGGCCGAGAACCTGATGATCTGACCTGTCCGCAAGCACTTC 1592
Qy      1621 AACAGGCTCTGGCGGATCGGCTCCGTCCAGTGCCTTCCAGGTGAGAGACTTAC 1680
Db      1593 AACAGGCTCTGGCGGATCGGCTCCGTCCAGTGCCTTCCAGGTGAGAGACTTAC 1652
Qy      1681 GCGCCCGTGCACGAGTGTGTGACACCTCCGGGAGACGCTGGGCGCCGACACAGACCC 1740
Db      1653 GGGCCCGTGCACGAGTGTGTGACACCTCCGGGAGACGCTGGGCGCCGACACAGACCC 1712
Qy      1741 GCGCCCGTGCACGAGTGTGTGACACCTCCGGGAGACGCTGGGCGCCGACACAGACCC 1800
Db      1713 GCGCCCGTGCACGAGTGTGTGACACCTCCGGGAGACGCTGGGCGCCGACACAGACCC 1772
Qy      1801 ACCATGAGATCATTCACGCGCTCGAGAGAGCGCCCGGGGCGCTTACTCCGGGCGCTC 1860
Db      1773 CCCATGAGATCATTCACGCGCTCGAGAGAGCGCCCGGGGCGCTTACTCCGGGCGCTC 1832
Qy      1861 GGATGGTTCGCTCCAGCGCGCGCGCGACCTCAGATCGTCACTCCGACCATCTGCTG 1920
Db      1833 GGATGGTTCGCTCCAGCGCGCGCGCGACCTCAGATCGTCACTCCGACCATCTGCTG 1892
Qy      1921 GCGGACGCGCGCGCGAGTTCGCGCTCGCGGCGCGCATCTGCTCCCTTCCGACCAAGAG 1980
Db      1893 GCGGACGCGCGCGCGAGTTCGCGCTCGCGGCGCGCATCTGCTCCCTTCCGACCAAGAG 1952
Qy      1981 GAGGAGTTCACCGAGACCGCTGTAAAGCGCCCGCCCTGCTCACCGCCCTTCCGACCGAGC 2040
Db      1953 GAGGAGTTCACCGAGACCGCTGTAAAGCGCCCGCCCTGCTCACCGCCCTTCCGACCGAGC 2012
Qy      2041 GCCGTGAGCGCGCGCGCATGA 2061
Db      2013 GCAGTGGCGGGCGCCAGCATGA 2033
```

```
RESULT 5
AAT58553
ID      AAT58553 standard; cDNA; 4496 BP.
AC      AAT58553;
DT      02-APR-1997 (first entry)
DE      Streptomyces pristinaespiralis snda and papa intergenic region.
XX
XX      Streptogramin B; antibiotic; biosynthesis; pristinamycin; virginiamycin;
KW      pipelicolic acid; cycloodeamination; papa; snda; snda; snda; snda;
KW      3-hydroxypicolinic acid; hydroxylation; ds.
XX
OS      Streptomyces pristinaespiralis.
XX
XX      Key      Location/Qualifiers
FH      CDS      /tag= a
FT      /product= "Snda"
FT      /note= "N-terminal coding region only, i.e. a partial
FT      open reading frame"
FT      607..1674
FT      /*tag= b
FT      /product= "Pipa"
FT      1800..2996
FT      /*tag= c
FT      /product= "SnbF"
FT      3018..4496
FT      /*tag= d
FT      /product= "Papa"
FT      /note= "N-terminal coding region only, i.e. a partial
FT      open reading frame"
XX
XX      MO9601901-A1.
XX      PD      25-JAN-1996.
XX
```

```
PF      04-JUL-1995; 95WO-FR000889.
XX
XX      08-JUL-1994; 94FR-00008478.
XX
XX      (RHON ) RHONE POULENC ROBER SA.
XX
XX      Blanc V, Thibaut D, Bamas-Jacques N, Blanche F, Crouzet J;
PI      Barriere E, Debussche L, Famechon A, Paris J, Dutric-Rosset G;
XX      WPI; 1996-097631/10.
DR      P-PSDB; AAW11584; AAW11585.
XX
XX      New streptogramin B derivs. useful as antibiotics - produced by new
PT      mutants of Streptomyces having altered genes for streptogramin B
PT      biosynthesis.
XX
XX      Example 1; Page 107-111; 146pp; French.
```

The papa gene of *S. pristinaespiralis* is involved in the biosynthesis of 4-dimethylamino-L-phenylalanine (DMPAPA), a precursor for pristinamycin IA. Upstream of the papa gene, on the complementary strand, is the snda gene coding for 3-hydroxypicolinic acid-AMP ligase. The region between these two genes was sequenced and two open reading frames were identified. The first (papa) decodes to an amino acid sequence with homology to ornithine cyclodeaminase from *Agrobacterium tumefaciens*. The pipA gene product is likely to catalyze the cyclo-deamination of lysine, leading to production of pipelicolic acid. Mutations in the pipA gene were shown to affect pipelicolic acid synthesis but not the synthesis of 3-hydroxypicolinic acid. The second open reading frame (snda) could be decoded to give a product with homology to hydroxylases of the cytochrome P450 type. Disruption of the pipA and snda genes can be used to produce strains of *S. pristinaespiralis* which are unable to produce the antibiotic pristinamycin I but which may be able to produce new, modified forms of it

Sequence 4496 BP; 659 A; 1980 C; 1321 G; 536 T; 0 U; 0 Other;

Query Match 22.3%; Score 459; DB 2; Length 4496;

Best Local Similarity 60.5%; Pred. No. 3; 6e-50; Matches 899; Conservative 0; Mismatches 495; Indels 93; Gaps 5;

```
Qy      2 TGCGCAGCGCTTCGATCGACCACTAGATCGTGTCCACCACTGTTCCAGTATCG 61
Db      3028 TCGGAACCTGCTGATCGACCACTAGATCGTGTCCACCACTGTTCCAGTATCG 3087
Qy      62 GCGAGGCCACCGGCGCAACCCCGCTGCTGCTGCCAAGCAGCGGCTGCGGCTGC 121
Db      3088 CCGAGGTGAACGGCGCGCGCTCGGCTGCTGCTGCGCAACGACACCGGACTGGCAGG 3147
Qy      122 CGCTCGAG-----GACTTGAAGCGATGCTGTGCTCCCGGCGCCCGGACCCGAC 175
Db      3148 CCTGCGCGCGCGCGCGCTGCGACCAAGTGTGTCTTCCACCGCGCCCGGCGCA 3207
Qy      176 GGGAAACGGAGCTTGGGAATCAGCGCGCGGCGGTCACGACGCGGCTGCGGCTCG 235
Db      3208 CCGACACCGACTTGGGCTTGAAGCGCGGCGGCGGTCACCGAATGGACCTTGGCTCG 3267
Qy      236 GCGTTCGCTCGGCGCACGAGGCGATGCCAGCTTTCGCGGAAACCGTGGGCTGCGCC 295
Db      3268 GGGTGTGCTGGGCGCACGAGGCGCTGTGCTGTGCGCGCGCGCGGCTGTCAAGCAC 3327
Qy      296 CCGAACCCATGACAGCGCGGCGGTCTCGAGGTGGCGACACCGGCGAGAGAGTCTTCGG 355
Db      3328 CCGAACCCATGACAGCGCGCGGCGGTCTCGAGGTGGCGACACCGGCGAGAGAGTCTTCGG 3387
Qy      356 GCGTCCCTCGCGGCTTACCGCGCGGTTCACCACTCGTGGCGCGGCGACCGACTCCCG 415
Db      3388 AATTCCTCTCCCGGCTGACCGTGTGCTTACCACTCGTGGCGCGGCGACCGACTCCCG 3447
Qy      416 ACGAGCTCGAACCCCTCGCTGAGGCGAGCGAGGCTGTCTTATGGGCTTCGCGACCGCG 475
Db      3448 CCGACTTGGCGCGACCGGCGCACACCGCGGAGCGGAGCGTGAATGGCGGTGCGCACCGCG 3507
```



```

QY 476 AGAAGCCGCTGTGGGCGTTCAGTTCCACCCGGAGTCCATCGCAGGCACTTCGGCCGG 535
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3508 ACCCTGCCCTTCGGGTGACAGTTCACCCCGAATGATCAGACAGGAACAGGCGACC 3567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 536 AGATCATGGCACTTCGGGACCT----- 560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3568 GGATGCTCGCAACTTCGGGACCTGCTCTGCGCGCGCGCGGACCGCCCCCGGACA 3627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 561 -----CGCCCTCGCCGACACACCGGGGACCGGGGACCGGGGCGGACTCCCG 605
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3628 CCGAAGCAATACCCGCAACCCGCAACCGCCCGCCCGCCCGCCCGGACCGGACCG 3687
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 606 -----GTACGAATCCACGTGCGCGCGCTGACAGTGTCCCGAGCG 646
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3688 CGTCCGCGCGCGGTGGGGAGTACCGGGCTGATGTGGCGAGGTGCGCTGTGCCGAGC 3747
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 647 CCGAAGAGTACGCCCGCGGTGCTGCGCGCGGAGGACACAGTTGTGGCTGAGACAGCA 706
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3748 CGGACGCGGCTTACCGGCTGTGCGCGGACCGCCCGGCTTGTGCTGACAGCA 3807
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 707 GCTCCGCTCTGAAAGCGGCTCGGCTTCTCTTCTCGGCGAGACCGGCGCGCGCTCG 766
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3808 GCGGCTCGAGCGCGGCTGCGCCGCTTACCTTCTCGGCGCGCCCGCGCGCGCTCG 3867
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 767 CCGAGTACTCTACCTACCGGCTGCGGACGCGCGGTCTTCGTCGCGGCTCGACGCA 826
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3868 GCGAAGATACCTACCTACGAGTGTGCGGACCGGCGCGGTGCGTCAAGAGAGGTTCA 3927
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 827 CCAAGACCGGAGGCGCGCGCTTCTTCACTGAGGAGGACAGCTCGAAGCGCGAC 886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3928 GCGAGACCGCGCGCGCGGACCTCTTGCAGCACCTGGAACAGAACTGG---CGGCC 3984
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 887 GGGTCCCGCTGCGCCCGAACTGCTTCTGAGTTCAACTCGGCTACGTGCGCTACTCG 946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3985 GCGGCTGCGCGCGCACCGGCTGCTTGTGAGTTCACTCGGCTACGTGCGCTACTCG 4044
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 947 GCTACGAGCTGAAGCGGAGACCAACCGGCGACCCCGCGACCGGTCCCGCACCCGACG 1006
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4045 GCTACGAGACCAAGCGCGACGCGGCGGCGAGGACCGCCACCGCGGAACTGCCGACG 4104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1007 CCGGCTTCTTCTGCGGACCGGCGCATGCGCTTCCAGCAACAGGAAGTGTGCTAC 1066
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4105 GCGCTTCACTGTTGCGCGACCGGATGCTGCGCTTCCAGCAACAGGCGGCGCTGCG 4164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1067 TGCTGCGCTCTGACCGCGCGGCGCACGACGCGCGCGCTGAGCTGCGGAGACG 1126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4165 TCGTGGACACTGACAGACACCGGACGCGCGCACCGACCGCGCGGAACTGAGCTCA 4224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1127 CCGAGACCTTACCGGCGCTGCGCGTCCGCGCGCGGCGGACCGGCGCGCATGCT 1186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4225 CCGACGCGCGCGGACCGCTGCGCACGCGCGCGCGCGCTTACCTGCTGCGCG 4284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1187 TCGGAGTCCCGGAGCGCGCGCGGCTTGGCCCCCGCGCGCGCGCACGACAAG 1246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4285 ACCACCAACTGCGCGCGCTGAGAGTCCATCAC-----CGCACAGGCTCG 4329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1247 ACAGCTTCACTCAAGCGCATGACGAGTGTCTCAAGAGATCCGCAACGCGGAGTGTGAC 1306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4330 CCGGCTACCGGGAATGCTGAGGAATGCGCGCGCTGATACCGGCGGAACTTACG 4389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1307 AGATGCTCTGACCAATGTCACCGCGCGGACCGAGGCGAGCGCGCTGCGCTTACT 1366
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4390 AGGTGCTGCTGAGCAATGCTCCGGGTGCGCGCGGAGTCACCGGCTACCGGCTTAC 4449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1367 CCGCGCTGGCGGCACTAGCGCGCTCGGACGCGCGCGCTGCTGAG 1413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4450 GCGGCTGGGCACTGACGCGCGCGCGCTTACCGGCTTACCTGAG 4496
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 6  
 AAT59268  
 ID AAT59268 standard; cDNA; 2888 BP.  
 XX

```

AC AAT59268;
XX
XX 02-APR-1997 (first entry)
XX
DE Streptomyces pristinaespiralis papA and papM intergenic region.
XX
XX Streptogramin B; antibiotic; biosynthesis; pristinamycin; virginiamycin;
KM DMPAP, precursor; 4-dimethylamino-L-phenylalanine; papA; papM; papB;
KM papC; isomerisation; aromatisation; N-methyltransferase; ds.
XX
OS Streptomyces pristinaespiralis.
XX
FH Key Location/Qualifiers
FH 1..687
FH CDS
FT /tag= a
FT /product= "papA"
FT /note="C-terminal coding region only, i.e. a partial
FT open reading frame".
FT CDS
FT /tag= b
FT /product= "papC"
FT 1873..2262
FT /tag= c
FT /product= "papB"
FT 2259..2888
FT /tag= d
FT /product= "papM"
FT /note="N-terminal coding region only, i.e. a partial
FT open reading frame"
XX
XX MO9601901-A1.
XX
XX 25-JAN-1996.
XX
XX 04-UTL-1995: 95WC-FR000889.
XX
XX 08-UTL-1994: 94FR-00008478.
XX
XX (RHON ) RHONE POULENC ROBER SA.
XX
PI Blanc V, Thibaut D, Bamas-Jacques N, Blanche F, Crouzet J,
PI Barriere J, Debussche L, Famechon A, Dutruc-Rosset G,
XX
XX WPI; 1996-097631/10.
XX
XX P-PSDB; AAW11582, AAW11583.
XX
XX New streptogramin B derivs. useful as antibiotics - produced by new
PT mutants of Streptomyces having altered genes for streptogramin B
PT biosynthesis.
XX
XX Example 1; Page 102-104; 146pp; French.
XX
XX The papA and papM genes of S.pristinaespiralis are involved in the
CC biosynthesis of 4-dimethylamino-L-phenylalanine (DMPAP), a precursor for
CC pristinamycin IA. The region between these two genes was sequenced and
CC two open reading frames were identified. The first (papC) was on the
CC complementary strand and decodes to an amino acid sequence with homology
CC to the region of E.coli TyrA which has been implicated in aromatisation
CC reactions. The papC gene product is likely to be involved in a similar
CC aromatisation of 4-deoxy 4-amino prephenate to give 4-amino
CC phenylpyruvate during DMPAP synthesis. The second open reading frame
CC (papB) could be decoded to give a product with homology to the region of
CC TyrA which has chorismate mutase activity. The papB gene product is
CC likely to catalyse the isomerisation of 4-deoxy 4-amino chorismate to 4-
CC deoxy 4-aminoprephenate. Disruption of the papB and papC genes can be
CC used to produce strains of S.pristinaespiralis which are unable to
CC produce the antibiotic pristinamycin I but which may be able to produce
CC new, modified forms of it
XX
XX Sequence 2888 BP; 390 A; 1319 C; 897 G; 282 T; 0 U; 0 Other;
SQ

```

Query Match 17.7%; Score 365.4; DB 2; Length 2888;  
 Best Local Similarity 74.2%; Pred. No. 3.1e-38;



	Matches	462;	Conservative	0;	Mismatches	161;	Indels	0;	Gaps	0;
Qy	1408	CTCGAGTTTCCCCGAACTGCGGTGTTGAGCGCCTTGCGCCGAGCGGTTTCTTACAGATCGGC								1467
Db	1	CTGCAAGTTCCTCCCGGGGACCAGTGTCTGACTGCTCTTCACTCCGAACGGTTCCTGCGATCGGC								60
Qy	1468	GCCGACGGCGGGGCGTGCAGTCCAAAGCCCATCAAGGGGAGCCCGCCCGGGGGGGGACACGGG								1527
Db	61	GCGAGCGCTTGGCGGAGTCCAAACCTCAAGGGGACCCCGCCCGGGGGCGCGGGCCCC								120
Qy	1528	GAGGAGACGAGCGGCTCCGGCGCCGACCTGGCGCGGCGGAGAAAGAACCGGGCGGAGAAC								1587
Db	121	GCCCAAGACCGCGCGCTCAAGGCTCTCCGCGCGGGCGGAGAAAGAACCGCAGCGAGAAC								180
Qy	1588	CTGATGATCTGTCGACTTGTGTCCGACGAACTTCAACAGCGTGTGCGGATGTGGCTCGTC								1647
Db	181	CTGATGATCTGTCGACTTGTGTCCGACGAACTTCCGCTCAGGTCTTGCAATCGGCTCGTTC								240
Qy	1648	CACGAGCCCGGGCTTCTGAGGTGAGACTAAGCGCCCGTGGCAACAGCTGTGTGACCC								1707
Db	241	CACGTAACGGGCTGTGTGAGGTGAGACTTACGCACTGTCACACAGCTGTGTGACGACG								300
Qy	1708	ATTCGGGAGACGGCTGCGGCGCCGAGACAGACACCGCGCTGCTGTAACGGCGGCTTCCG								1767
Db	301	GTCGGCGGCGCGCTGTGGCGCGGACGTCCTCCGCGCCCGCGGGGAGTACGGGCGGCTTCCC								360
Qy	1768	GCGCGGCTTCATGACCGCGCGCGCCCAAGAGCGCACATGAGATCATGACCGCTGGAG								1827
Db	361	GCGCGGCTTCATGACCGCGCGCGCCCAAGAGTCCGACCATGAGTTCATGACCGGCTGGAG								420
Qy	1828	GAAAGCCCCCGGGGCGTCTACTTCCGGGGCGCTCGGATGTTTGGCTTACGAGCGGCGCC								1887
Db	421	AAGGACCCCGCGGCGGTGTACTTGGGCGCGCTGTGGCTTACTTCCGCTTACGAGCGGCGCC								480
Qy	1888	GACCTCAAGCATCTGTCATCCGACACCATCGTGTGCGCGGACGAGAGAGTTCAGGCGGTC								1947
Db	481	GACCTCAAGCATCTGTCATCCGACACCATCTGCGCACGAGAGAGCGCCACCATCGGCGTC								540
Qy	1948	GCGCGGCGGATCGTGTCTCTCTCCGACCGAGAGAGAGTTCACCGAGACCGTGTAAAG								2007
Db	541	GCGCGGCGGCGTGTGTGCGCTGTCCGACCCGACGAGCGATTCGCGAAATGCTCTCAAG								600
Qy	2008	GCCGCGCGCATGATCACCAGCCCT								2030
Db	601	GCGCAGACCACTTCGCGCGCT								623
RESULT 7										
ADN36903	ID	ADN36903	standard; DNA; 2151 BP.							
AC	XX	ADN36903;								
DT	XX	15-JUN-2004	(first entry)							
DE	XX	X. albilineans XALBI gene cluster DNA encoding protein AlbXVII.								
KM	XX	AlbXVII; albicidin biosynthetic gene cluster; biosynthetic gene cluster;								
KW	XX	AlbXVII; albicidin biosynthetic gene cluster 1; phytotoxic damage; AlbXVII;								
KW	XX	gene; ds.								
OS	XX	Xanthomonas albilineans.								
PN	XX	WO2004035760-A2.								
PD	XX	29-APR-2004.								
PF	XX	17-OCT-2003; 2003WO-US033142.								
PR	XX	18-OCT-2002; 2002US-0419463P.								
RA	XX	(UYFL ) UNIV FLORIDA.								
RA	XX	(CIRA-) CIRAD CENT COOP INT EN RECH AGRONOMIQUE.								

PI	Royer M, Gabriel DW, Frutos R, Rott P,	
XX	WP1; 2004-365158/34.	
DR	P-PSDB; ADN36934.	
XX		
PT	New transformed host cell, useful for producing antibiotics, preferably	
PT	polysaccharide antibiotics for protecting plants against phytotoxic damage,	
PT	or damage against abdicidin.	
XX		
PS	Claim 8; SEQ ID NO 11; 193bp; English.	
XX		
CC	The present invention relates to a novel abdicidin family of antibiotics	
CC	produced by the expression of biosynthetic gene clusters from xanthomonas	
CC	albilineans designated as XALB1, XALB2 and XALB3 (abdicidin biosynthetic	
CC	gene clusters 1, 2 and 3). The invention discloses the polynucleotide	
CC	sequences of these gene clusters, and the proteins encoded by the open	
CC	reading frames (ORFs) within the gene clusters. Also disclosed are	
CC	methods for producing an antibiotic and protecting a plant against damage	
CC	from abdicidin and against phytotoxic damage. The present sequence	
CC	represents an ORF from the xanthomonas albilineans XALB1 gene cluster.	
XX		
SO	Sequence 2151 BP; 418 A; 607 C; 665 G; 461 T; 0 U; 0 Other;	
	Query Match 17.6%; Score 361.8; DB 12; Length 2151;	
	Best Local Similarity 51.4%; Pred. No. 9.3e-38;	
	Matches 1057; Conservative 0; Mismatches 932; Indels 66; Gaps 7	
QY	2 TGCGCAGCGCTTCGATGAGCAACTGACACTGCTTCCAGTACATCG 61	
DB	2 TGGCGTCCCTTATCATTAATATATTAACGATTCCTGACCTTGAATTCGCCGACTAGCTAG 61	
QY	62 GCGAGGCCACCGGGGCAACCCCGCTGCTGCGCCCAACGACGCGGACTGTGCGG----- 116	
DB	62 CGGAGATCTTCGGGGAAGATCCCTGTGGTGGCAACAAGAGATACCTCTGGCACGAAC 121	
QY	117 -GCTGCCCGTGAAGACTTGAACGCGATCGTGTGTCCCGGGCCCGGCAAGCCCGAC 175	
DB	122 TGAAGGACCCCGGGGGAATTTCTCGATCATCTGTTCCCGGCTCCGGCTCGGTGTGA 181	
QY	176 GGGAGCGGACTTCGGAATCAGCCGCGGGGAGATCACCGAAGGGCCCTGCCGCTCTCG 235	
DB	182 ATGAGCGGAAATTTACATCTCGTGAAGGGCGCTGAGACAAAGAAATTTCCGGTGTAG 241	
QY	236 GCGCTGCTCTCGGCAACAGGAGCATCGCCAGCTCTTCGCGGGAACGTCGGGCTCGCCC 295	
DB	242 GCGATATGCTGGGCTTCAAGGACTTGGCATGTCTATGTGTGGCCGATCTCGATATGCG 301	
QY	296 CGGAACCATGACCGCGCGGAGTTCGAGAGTGGCGGACACCGGCGAAGACGTTCCTCGGG 355	
DB	302 CGGTGCCCTTCAATGGCGCTCGCTCCACCGCATCAACACCGGCGACGGTGTGTTGGAAG 361	
QY	356 GCGTCCCTCGCGCTTACACCGCGGTGGCTTACACTCTCCCTGGCCGACCGA-----CG 409	
DB	362 GCATCCCGGACCGCTTCGAGGCAAGTGGCTATCATCTGTTAGTGTCTGCGAGCAATGCG 421	
QY	410 TCCCGGACGAGCTCGAACCCCTCGCTGAGAGCGACGAGGGGTGTCATGGAGCTGGCGGC 469	
DB	422 TGGCGCTGTGTCTGAAGTGAACGCGCGGTACCGATTGGCGTGTGTGATGGAGCTTGGAC 481	
QY	470 ACCGCGAAGACCGCTGTGGGGCGTCCAGTTTCCACCGGAGTTCATCGGACGCACTTTCG 529	
DB	482 ACGTGAACAACCGGAATGGGGAGTACAGTTTCCACCCCGAATGATCTCTCAACGAACG 541	
QY	530 GCGGGAAGATATGGCAATCTTCGCGACCTCGCCCTCGCCACACCGGGGACGCGGCC 589	
DB	542 GCAAGGCGCATTTGTGTCTTAACTTTTCCCAAGCTGTGCTGCGCGCACAGTGCACCTTTACTTG 601	
QY	590 ACGGGGCGCATCCCGGTACGAATCCACGCTGGCGCGCGGTGAGTGTGCGGACGCGCG 649	
DB	602 CCGGCTGGAGCAGCGCGGCAAGGTTTAAAGCTTTTGGCGCCCGGAAGTGTGACACGCC 661	
QY	650 AAGAGTACCGCGCGCTGCTGCCCGGCGAGGCAACAGTTTCTGCTGGA---CAGCA 706	



```

Db      662 GGGTAGCTGCGATGAGCGCGGAAGATCAATGCGCGTTGGCAGCGGAAGATGCTTTTC 721
QY      707 GCTCCGCTCTGGAAGGCGCTCGCTTCTCTCTCTCGCGAGC-----GACCGGCGCC 760
Db      722 TGCGCTTGTGCTGAGGAAAGCATTTGCTTGGCTGAGCAGCCAGCTGCTTGCAATC 781
QY      761 CGCTCCCGGAGTACTACCTTACCGCGTCCGCGAGCGCTGCTCCGCTCGCGAGCTCG 820
Db      782 CAATGGCGCGCTTATGCTTATGAGAGCGGTGAACGAGCGAGAGTATGCGGCTTGGC 841
QY      821 ACGGACACAGACCCGAGCGCGCGCTTCTTCACTAAGAGGAGGAGGAGGAGGAGGAG 869
Db      842 TGGGCGAGGAGCATGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 901
QY      870 -GAGCTCGAAGCCCGAGCGGTCCCGTCCGCGCGCGAGCTGCGCTTCAAGTTCAACTCG 928
Db      902 CGTTGCAATCGTGTCTTACTGAGGAGCGTCCGAGCGCGCAGCGTTCGCTTTCGCGCG 961
QY      929 GCTACGTCGCTACTCTGCGCTACGAGCTGAGGCGGAGACCAACCGGCGACCCCGCGCAC 988
Db      962 GCTACGTCGCTACTGAGTACGAGTACGAGTACGAGTACGAGTACGAGTACGAGTAC 1021
QY      989 GGTCCCGCAGCCCGAGCGCGCGCTTCTTTCGCGAGCGCGCGCATGCGCTGACCAAC 1048
Db      1022 CCAATGCGATCCCGATGCGCTTGTGATGCGCGTGGAGCGCTTGTGCTTGCAGCAGC 1081
QY      1049 AGGAGGCTGCTGCTACTGCTGCTGCGCTTGCAGCGCGGCGCGACGAGCGCGCGCG 1108
Db      1082 CCACTGAGGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1133
QY      1109 CTGCTGCTGCGGAGAGCGGCGGAGACCTTCAACCGGCTGCGCGCGCGCGCGCGCGAGC 1168
Db      1134 ATTGCTTGTGCTGAGACCGCATCGAGCAGCATATGCTGCTGCTGCTGCTGCTGCTGCT 1193
QY      1169 CGACCCCGCGCATGCTTTCGGAATCCCGAGCGCGCGCGCTTTCGCGCGCGCGCGCG 1228
Db      1194 TTGCATTTGCTGAGCGCTGCGCGAGCATGAGAAATGAGACTCAATCATGATGCTC 1246
QY      1229 GCGCGCGCGCGACGAGAGAGCGCTTACTTCAAGCGCATGAGCGAGTGTCTTCAAGAGATCC 1288
Db      1247 -----GCGGCTACCTTGAAGCAATCGAGCGTTCGAAACCAACGCAATCG 1288
QY      1289 GCAAGCGCGAGTGTGAGATGCTGCTGACCAACATGATGATCGCGCGCGCGAGGAGCA 1348
Db      1289 TCGATGCGCGAGTCTTATGAAATCTGTTCACCACTGCTTCTGCTTCAAGCGCGAGCTG 1348
QY      1349 CGGCGCTGCGCTTACTCTCGCGCTGCGCGCATCAGCCCGCTGCTGCTGCTGCTGCTGCT 1408
Db      1349 ATCATTTGATGCTCTATCGCTACATGCGGAGGAGAACCCAGCGCGCTTTCGCGCGCTAT 1408
QY      1409 TCGAGTTCCCGGAATCTGCTGCTGAGCGCTTGCAGCGCGGCTTCTTCAAGATCGCGCG 1468
Db      1409 TGGGTAAAGGTACGATTTATCTTAACTTACAGAGCGCTTCTTCAAGATGAGAGC 1468
QY      1469 CCAGCGCGCGCTGCTGAGTCCAGGCCATCAAGAGGAGAGCGCGCGCGCGCGCGAGCGCG 1528
Db      1469 GCCACGCGCAGATTCAGACCAAGCCATCAAGGAGCCTTCCGCGCGCGAGAGATCCCG 1528
QY      1529 AGGAGAGCAGAGCGGCTCGCGCGAGCTGCGCGCGCGAGAGAGAGCGCGCGAGAAC 1588
Db      1529 AACTGAGACCGTAACTTGGCATGGCGCTGCGCGCTGCGAGAAAGAGCGAGCGGAAACT 1588
QY      1589 TGAATGCTGCTGAGCTGCTGCTGCGCAAGCACTCAAGAGCTGCTGCGAGTGGCTCGCTC 1648
Db      1589 TGAATGCTGCTGAGCTGCTGCTGCGCAAGCACTCAAGAGCTGCTGCGAGTGGCTCGCTC 1648
QY      1649 ACGTGGCGCGGCTTGTGAGGTGAGAGCTTACGCGCGCGGAGCGCGAGCTGCTGAGCA 1708
Db      1649 CCGTGGCGCAGCTGATGAGCATGAGAGCTTCAAGAGCGTGCATCATGATGCTGAGCA 1708
QY      1709 TCCGCGGAGCGCTGCGCGCGCGAGCAGCAGCGCGCGCTGCTGAGCGCGCGCTTCCCGC 1768

```

```

Db      1709 TGGAGCGAGGCTCGCGCGCGATTCGAGTCTAGTCACTGCTTAAAGCGGTCTCCCG 1768
QY      1769 GCGGCTTCATGACCGCGCGCGCGCAAGAGCGGACCAATGAGATCATGACCGCGCTGAGG 1828
Db      1769 GCGGCTTCATGACCGCGCGCGCGCAAGAGTGGAGATGAGATTTATGATGCTGAGAG 1828
QY      1829 AAGCGCGCGCGGCTCTACTCCGCGCGCTCGAGTGTTCGCTTCAAGCGCGCGCGCG 1888
Db      1829 ATGCGCGCGGCTGCTGCTTATGCGAGCATGCGCTACCTGGGCTTCACTGCTGCGCG 1888
QY      1889 ACTTACGATCTGCTATCCGACCATGCTGCTGCGAGCGGAGCGGAGCTTGGCTG 1948
Db      1889 ACCTAAACATTTGCGATCCGAGCTTCTTATGACCGGAGGAGAAATACGTTTCGCGCG 1948
QY      1949 GCGGCGCGATTCGCTCTCTCCGACGAGAGGAGGAGTTCACCGAGACCGTGTAAAG 2008
Db      1949 GCGGCGCGATTCGCTCTCTCTCCGACGAGAGGAGGAGTTCACCGAGACCGTGTAAAG 2008
QY      2009 CCCGCGCATGTGCA 2023
Db      2009 CCGAGCGCATTCCTCA 2023

RESULT 8
ADN36893
ID      ADN36893 standard; DNA; 55839 BP.
XX
AC      ADN36893;
XX
DT      15-JUL-2004 (first entry)
XX
DE      X. albilineas XALB1 gene cluster DNA.
XX
KW      Albicidin family; antibiotic production; biosynthetic gene cluster;
KW      XALB1; albicidin biosynthetic gene cluster 1; phytotoxic damage; gene;
KW      de.
XX
OS      Xanthomonas albilineas.
XX
PN      WO2004035760-A2.
XX
PD      29-APR-2004.
XX
PF      17-OCT-2003; 2003WO-US033142.
XX
PR      18-OCT-2002; 2002US-0419463P.
XX
PA      (UYFL ) UNIV FLORIDA.
PA      (CIRA-) CIRAD CENT COOP INT EN RECH AGRONOMIQUE.
XX
PI      Royer M, Gabriel DW, Frutos R, Rott P;
XX
DR      WPI; 2004-365158/34.
XX
PT      New transformed host cell, useful for producing antibiotics, preferably
PT      polyketide antibiotic for protecting plants against phytotoxic damage,
PT      or damage against albicidin.
XX
PS      Claim 8; SEQ ID NO 1; 193bp; English.
XX
CC      The present invention relates to a novel Albicidin family of antibiotics
CC      produced by the expression of biosynthetic gene clusters from Xanthomonas
CC      albilineas designated as XALB1, XALB2 and XALB3 (albicidin biosynthetic
CC      gene clusters 1, 2 and 3). The invention discloses the polynucleotide
CC      sequences of these gene clusters, and the proteins encoded by the open
CC      reading frames (ORFs) within the gene clusters. Also disclosed are
CC      methods for producing an antibiotic and protecting a plant against damage
CC      from albicidin and against phytotoxic damage. The present sequence
CC      represents Xanthomonas albilineas XALB1 gene cluster.
XX
SQ      Sequence 55839 BP; 10970 A; 16145 C; 17705 G; 11019 T; 0 U; 0 Other;
Query Match      17.6%; Score 361.8; DB 12; Length 55839;

```



Best Local Similarity 51.4%; Pred. No. 4.8e-38;  
Matches 1057; Conservative 0; Mismatches 932; Indels 66; Gaps 7;

```
QY      2  TGCGCAGCTTCTGATCGACAACTACGACTGTTACCCACAACTGTTCCAGTACATCG 61
      14910 TGGCTGCTGCTTATCATTAATATTAATTAATGTTTCACTTGGAATTCGCCACTATGAG 14969
QY      62  GCGAGGCCACCGGGCAACCCCCGCTGCTGTCGCCAAAGAGCCGACTGTGCGC----- 116
      14970 CGGAGATCTTCGGGAGAGATCCCTGGTGGTGCACAAAGAGATTAATCTCTGGACAGAAC 15029
QY      117  -GCTGCCCTTGAGAGACTTCAGACCGATGCTGTGTCCCGGGCCCGGACGCCCAAC 175
      15030 TGAAGGACCGCGGGGATTTTCTCTGATCACTGTTGCGCCGCGCTCCGCTCGGTGTTA 15089
QY      176  GGGAAAGGGGACTTGGAATACGCGCGGGGATCACCGAAGGGGCTGCGCTCTCG 235
      15090 ATGAAGGAGATTTTCACTCTCGTGCAGGCGCTGAGCAGAAAGAAATTTCCGGTGTAG 15149
QY      236  GCGTCTGCTCTGCGCACAGGGGCAATCGCCAGCTCTTGCGGCAACCGTCCGCTCGCC 295
      15150 GCGTATGCTGCGGCTTTCAAGGACTTGGCATGTCTATGTTGGCCGATCTGTGATGCG 15209
QY      296  CCGAACCATGACAGCGCCGGGTCTCGAGGTGCGGCAACCGGCGAGAGACGTCTTCGAG 355
      15210 CCGTGCCTTTCATGCGCGCTGCTCAACCGTCAACACCGGCGACGGTTGTTGAG 15269
QY      356  GCGTCCCTCGCGCTTACCGCGCGTGGCTACCACTCCCTGCGCGCAACGA-----CC 409
      15270 GCAATCCGCGAGCGTTTGAAGGCAATGCGTATCACTGTTGATGCTTGCACGAAATCG 15329
QY      410  TCCCGCAGAGCTCGAACCCCTGCTGAGCGCAGCAGGGGCTGATGGGCTGCGGC 469
      15330 TGGCGCTGTGCTGAAGTACGCGCGCTACCGATTGCGGTGTGATGAGCTTGGAGC 15389
QY      470  ACCGCGAAGACCGCTGTGGGCGTCCAGTTCCACCGGAGTCCATCGGAGCACTTCG 529
      15390 ACGTGAACACCGGAAATGGGAGTACAGTTCCACCCGAAATGATCTTCAACGAAACAG 15449
QY      530  GCGCGGAGATCATAGGCCAATTCGCGGCACTCGCGCTCGCGCCACACCGGCGACGGCGC 589
      15450 GCAAGCGCATTTGTCTACCTTTGCTTGCACAAAGCTGTGCGCGCACAGTGCACCGTTACTTG 15509
QY      590  ACGGGGCGGACTCCCGCTAGCAACTCCACGAGGCGCGGCTGACGTGCTGCGGACCGC 649
      15510 CCGGCTGAGACAGCGCGCGGAGGTTTAAAGCTTTGCGCGCCGAGATGTGACACCGC 15569
QY      650  AAGAGTACCGCGCGGCTGCTGCGCGCGAGGAGCAACGTTCTGCTGGA---CAGCA 706
      15570 GGGTACGTGCGATGTGAGCGCGGAAAGTCAAGTCCGTTGGCAGGCGGAAATGTCTTTC 15629
QY      707  GCTTCGCTCTGAAGGCGCTCGCGCTTCTCTTCTGCGGCA-----GACCGCGCGC 760
      15630 TGGCTTGTTCGCTGACGAAACATGCTTCTGCTGACAGCAGCTGATGATGAGTGC 15689
QY      761  CGCTCGCGAGTACCTCACTACCGGCTGCGCGACGCGCTGTCTCGCTCGCGCTCGC 820
      15690 CAATGGGCGGCTATTGTTTATGAGAGCGGTGAAAGAGACGAGAGTGCAGCATTTGCG 15749
QY      821  ACGGCAACAGACCGGAGCGCGCGCTTCTTTCACACTACCTGAGAGA-----869
      15750 TGGGCGCAGGAGACATGTGTGACAGAGGACGAGCGAGGTTTCTTGTGATGATGATCGG 15809
QY      870  -GCACTCGAACCGCGAGCGGTCCCGTGCAGCGCGAACTGCGCTTGAAGTTCACTCG 928
      15810 CGTTGCAATCGGTCTTACTGAGAGAGTCCCGAGCGGCGCACCGTTTCGCGTTCGCGGCG 15869
QY      929  GCTACGTCGCTACCTCGGCTTACGAGTGAAGCGGAGACCAACCGCGAGACCCCGCGCAC 988
      15870 GCTACGTCGCTACGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 15929
QY      989  GGTCCCGCAACCGGAGCGCGGCTTCTTCTGCGCGAGCGGCGATGCGCTGACACAC 1048
```

```
Db      15930 CCAATGCCATCCCGATGCGTTGTGATGCGCGCTGAGAGCGCTTGTGCTTCGACACAG 15989
QY      1049 AGGAAGGCTGCTGTACTCTGCTGCGCTGACCGCGCGGGGCGACAGAGAGGCGCGCG 1108
      15990 CCACTGAGAGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 16041
QY      1109 CTTGCTGCGGAGAGAGCGGCGGAGACCTTACCGGCTGCGCGCTGCGCGCGCGCGAGC 1168
      16042 ATTGCTGTGCTGATGACCGCATGAGAGACCTATTCATGCTATGTGTACGCGCTCGGC 16101
QY      1169 GACACCCCGCATGTGTCTTGTGGATCCCGAGCGCGCGGCGCTTGGGCTTCGCGCC 1228
      16102 TTGCATTTTCCGTAGGCTGCGGAGAGTGAATTCGAGCTCAATCATGTGTGTC----- 16154
QY      1229 GCGCGCGCCACAGAGAGAGCTTACTTCAAGCGCATTCGACGAGTGTCTTCAAGAGATTC 1288
      16155 -----GCGGCTTACCTTGAAGCAATTCAGCGCTTGCMAACACGCAATCG 16196
Db
QY      1289 GCAACGGCGAGTGTGATCGAGATCTGCTGAGCACAATGTGACCGGCGCGAGCGGAG 1348
      16197 TCGATGCGGAGTCTTATGAATCTGTCTTACCGACTGTTCTGTCTCAGGCGGAGCTGG 16256
QY      1349 CCGCCTTGCCTCTACTCTCGCGCTGCGCGGCTCATCAGCCCGCTCCGTAAGCGCGCTGC 1408
      16257 ATCATTTGATGCTCTATTCGTACATGCGGCGAGAGAACCAAGCGCGCTTCGCGGCTATT 16316
QY      1409 TCGATTTCCCGAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1468
      16317 TCGTAACTGTAAGTATGATCTTATCTTAACTTCAACAGAGCGTTTCTGGAATGAGAGC 16376
QY      1469 CGACGCGCGCGCTGATCGATCCAGACCATCAAGGGGAGACCGCGCGCGGCGGAGCGCG 1528
      16377 GCCACGCGAGATTCAGACCAACAGCAATCAAGGGGAGACCTGCGCGCTGCGAGAGATCCC 16436
QY      1529 AGGAGAGCAGCGGCTTCGCGCGCGAGCTGCGCGCGCGCGCGGAGAGAGACCGGCGGAGAAC 1588
      16437 AACTGACCGTAACTTGGCATGCGCGCTGCGCGCTTGGGAAAGAGACCGAGCGGAAACT 16496
Db
QY      1589 TGAATGCTGCTGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1648
      16497 TGAATGCTGCTGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 16556
QY      1649 ACGTCCCGCGGCTCTTGAAGTGAAGCTTACCGCGCGCGTGAACAGCTGTGTGTGTGTGTGT 1708
      16557 CCGTCCCAAGCTGATGACATTCGAAAGCTTACAAAGCGTGTATCAGATGTGTGTGTGTGT 16616
Db
QY      1709 TCGGAGGAGCGGCTGCGCGCGCGGACACGAGACCGCGCTGCGTACCGCGCGCTTCCCG 1768
      16617 TGAAGCGAGGCTGTGCGCGCGCGGATGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 16676
QY      1769 GCGGCTTCATGACCGGCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1828
      16677 GCGGCTTCATGACCGGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 16736
Db
QY      1829 AAGGCGCGCGGCGGCTGTCTCTCGGCGCGCTGAGATGTGTGTGTGTGTGTGTGTGTGTGTGT 1888
      16737 ATGCGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 16796
QY      1889 ACCTCAGATTCGATTCGCGCACATCGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1948
      16797 ACCTTAACTTGTGATTCGCGAGCTTCTTATGACGCGGAGAGAAATACGTTTGTGCGCGCG 16856
QY      1949 GCGGCGGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2008
      16857 GCGGCGGATTCACCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 16916
Db
QY      2009 CCGGCGCATGTGTCA 2023
      16917 CCGAGCGATCTTCA 16931
```

RESULT 9  
AAS59535



ID AAS59535 standard; DNA; 26309 BP.  
XX AAS59535;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Propionibacterium acnes immunogenic protein encoding DNA #30.  
XX  
KM SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KM dermatological; osteopathic; neuroprotectant; ds.  
XX  
OS Propionibacterium acnes.  
PN M0200181581-A2.  
PD 01-NOV-2001.  
PF 20-APR-2001; 2001MO-US012865.  
PR 21-APR-2000; 2000US-0199047P.  
PR 02-JUN-2000; 2000US-0208841P.  
PR 07-JUL-2000; 2000US-0216747P.  
XX  
XX (CORI-) CORIXA CORP.  
PI Sheiky YMW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX WPI; 2001-616774/71.  
XX  
PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris.  
XX  
XX Claim 1; SEQ ID NO 30; 1069bp; English.  
XX  
CC Sequences AAS59506-AAS59804 represent DNA molecules encoding  
CC Propionibacterium acnes immunogenic polypeptides. The proteins and their  
CC associated DNA sequences are used in the treatment, prevention and  
CC diagnosis of medical conditions caused by P. acnes. The disorders include  
CC SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and  
CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in  
CC infections of bone, joints and the central nervous system, however it is  
CC particularly involved in the inflammatory lesions associated with acne  
CC vulgaris. A method for detecting the presence or absence of P. acnes in a  
CC patient comprises contacting a sample with a binding agent that binds to  
CC the proteins of the invention and determining the amount of bound protein  
CC in the sample. The polypeptides may be used as antigens in the production  
CC of antibodies specific for P. acnes proteins. These antibodies can be  
CC used to downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the  
CC polypeptide shown in A4U6704-A4U6985 and A4U67509. Note: The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 26309 BP; 5546 A; 8173 C; 7885 G; 4699 T; 0 U; 6 Other;  
Query Match 15.3%; Score 316; DB 4; Length 26309;  
Best Local Similarity 50.6%; Pred. No. 3.7e-32;  
Matches 1057; Conservative 0; Mismatches 955; Indels 78; Gaps 9;  
QY 10 CTTCTGATCGACACTGCTGTCACCCCAACTGTTCCAGTACATCGGAGGCC 69  
DB 19015 CTGCTGTTGATCAACACGATTCCTTCACTTCACTGGACACCTATTGGCCGAGATC 19074  
QY 70 ACCGGGCAACCCCGCTGTCGTGTCACGACGCGCAGTGTGCG----- 114  
DB 19075 AGCGGATCGAGCGAGTGTGTGTCGCGCCGAGAGAGTTCAGAGGTGCGAGGATCTCTCGC 19134

QY 115 CGGCTGCCCGTGAGAGCTTTCGACGCGATGTCGTGTCCCGGAGCCCGGACGCCCGAC 174  
DB 19135 CGATTGGCTGATGGCGAATTTCGACCGATGTTATGATTCAGAGACTTGCTCCCGGAC 19194  
QY 175 CGGGAACGGGACTTGGGAATTCAGCCCGCGGCGATCA---CGACAGCGGCTGCCCGTC 231  
DB 19195 AACGACAAAGACTTTCAGAGACCGCGGACAGGTGATGATGCTCCAAAGCTCCCGCTG 19254  
QY 232 CTCGGGTCGTCGCTCGGCGACGAGGACATGCGCCACTTCTGGGCGGAACCGTCGCGCTC 291  
DB 19255 CTCGGGTCGTCGCTCGGCGACGAGGCTCG--GGTTACGTACGCTGCCAGTTTGACG 19311  
QY 292 GCCCGGAACCCATGACGCGCGGCTCCAGGTGCGGACACCGCGGACGACGCTTTC 351  
DB 19312 ACCATTGCGGCCCATATCAGCGCATGCTGACCGCAATTATCATCTCCAGGCGTGAATCTTC 19371  
QY 352 CGGGGCTTCCTCGCGCTTTCACCGCGGTGCTACACTCCCTGCGCGCACCGACCTTC 411  
DB 19372 GCGGATTAACCGCAAGATTTTCGAGGCACTCGGTATCATCTGTGTTGACACGTTTG 19431  
QY 412 CCGAGAGACTGGAACCCCTCGCTGAGAGACGACGAGGCTGTATGGGCTTCGCGCAC 471  
DB 19432 GGTGACCAATGCTCGAATGCTGCGCGCGGAGATGGGCGCATGCGCTTCGAGGTT 19491  
QY 472 CGGAGAGCGCGCTGTGGGCGCTCCAGTTCCACCCCGAGTTCATCGGCGACGACTTCGCG 531  
DB 19492 GCGGACCGGCGACATCGGGGCGGTGCAATTCACCCAGATCACTATGACGAGTTCGGA 19551  
QY 532 CGGAGATCATGGCCAACTTCGCGACCTCGCCCTTCGCCACACCGGGGACGCGCGCAC 591  
DB 19552 TGTCAATTATGACGAATTTCTCGGCGATCGGAGCGACGACACACAGCAGGTACCG 19611  
QY 592 GGGGCGGACTCCCGCTGACGACTTC-----AGTGGCGCGCTGACGTCGTGCGG 643  
DB 19612 GACMAACGGGAACCGGCAACATTCAGTCAGAAATGTGCGCGGACTCGGCGCGGTG 19671  
QY 644 AGCGGAGAGGTACGCGCGG-----CTGCTGCGCGGCGGAGGACCAAGTTCTGGCTG 699  
DB 19672 ACATATCACACCAACAGGTGACATGACCTTCAGGAAGAACCACTTCAGCGGCTTC 19731  
QY 700 GACAGAGCTCGCTCTTCGAAAGCGCGCTTCCTTCCTTCGCGGACGACGCGGCG 759  
DB 19732 GCTGGGAGAGGTGAGCGCTTGTGTGACTCGGCGCACACGCGTGGCGACACCGGCGGT 19791  
QY 760 CGGCTGCGGAGTACTCTACCTTACCGCGCTGCGGACGCGGTGTCTCGCTCGCGGCTCG 819  
DB 19792 TGGAGTGTATGGGCACGGCGTTCGCGGCGGCGGATCCGAATGCTCATGACGTCATC 19851  
QY 820 GACGAGCACACGACCGGACCGGCGCGCTTCCTTCAACTACCTGAGAGGACGTCGAA 879  
DB 19852 ACAAAACGTTGACGGTCAACGCGGCAACCACTGTGTGATGTGCTGGACTTTTGGAA 19911  
QY 880 CGCGGACGGGTCCCGCTGCGCCCGCAACTGCGCTTCGA-----GTTCAAC 924  
DB 19912 CGCGCTTTTGCAATGAGATGATACGAGAGGACCGGACCTGATGATCACTCCATTCAACA 19971  
QY 925 CTCGCTTACGTGGCTTACTCTCGGCTTACGAGCTGAGGCGGAGACACCGGCGACCGCGCG 984  
DB 19972 GGTGCTTATGTGCGTTTCTCGGCTTATGATGCAAGCGCTGACTTGGGCGCGGAATGG 20031  
QY 985 CACCGGTCCCGGACCGCGGACCGCGGCTTCCTTCGCGGACCGCGGCGCATCGCGCTCGAC 1044  
DB 20032 CACTGAGAGAAATGCGCGATGCAATGATGATGCGTTCGCGGACGCTGATGCTGATCGAC 20091  
QY 1045 CACCAAGAGGCTGCTGCTTACTGCTGCGCTTCGACCGCGGGGCGACACGACGCGCGCC 1104  
DB 20092 CACTTACCGGATCAAGCTTACCTTCTAGCTTTGAGACGACGCGCGGCTGCAAT 20151  
QY 1105 CGGCTGTGCTGCGGAGACGCGCGGACGACCTTCACCGGCTGCGCTCGCGCGCGCGCC 1164  
DB 20152 GAGCGCGTGTATC-----TATTGAGACGACTGCTGCGACGCTGCTG 20192



QY 1165 GAGCCGACCCCGCATGTTCTTCGGATCCCGGAGCGCGCGCTTCGCCCCCTG 1224  
 DB 20193 GAAGCCGACGATGAGCAGCAAGAGCCCGACCGCTAACCGTGTCTTGGAAGGCTGTG 20252  
 QY 1225 GCGCGCGCGCCACGACCAAGAGCCTTACTCAAGCGCATCGAGATGCTCAAGAG 1284  
 DB 20253 -----GGGGCTCAGCCGAGGGGGCTACAGGACCGTGTGCTGCAATTCAGCAGGCG 20304  
 QY 1285 ATCCGCAACGCGAGTGTGATGAGATCTGCTGACCAATGATGATCAGCGCGGACCGGAC 1344  
 DB 20305 CTAAACGCGCGGTGATCTGTACAGAGCGCTGTCTGACGACACGTGACCTCCCGGTGAC 20364  
 QY 1345 GCGACGCGCCCTGCGCTCTACTCCGCGCTGCGCGCCATCAAGCCCGTCCGTAAGCGCC 1404  
 DB 20365 GTGGAAGCGCTGGGGTGTGTACCGCGCATCCGGGGTGGAAACCAAGCCCTTATGGGGCG 20424  
 QY 1405 CTGCTCGAGTTCCCGCACTG---TCGGTGTGAGCGCTTCGCGCGAGCGGTTCCTACG 1461  
 DB 20425 TATCTTCGTTTACCGACCGCGGTGTGAGAGGTGTCTTTCATCCCGGAACGGTTCTGTG 20484  
 QY 1462 ATGCGCGCGACGCGCGCGGTGAGTCCAGCCCATCAAGGGAGACCCCGCGCGCGCG 1521  
 DB 20485 CGAGTACGCAACGCGCATGCTGATTCAGCCCGATCAAGGGAGCAATGCTGCTGTGAT 20544  
 QY 1522 ACCGCGAGGAGAGCAGCAGCGGCTCCGCGCGCATCTGCGCGCGCGGAGAAAGAACCGGCGC 1581  
 DB 20545 GATCCCGTGAAGATGCCCCGGGAATTGTGATTTACAAATGACCGGAAAACCGCGCC 20604  
 QY 1582 GAGAACCTGATGATCTGTCGACCTGTGTCGCAACGACCTCAACAGCGTCTCGCGCATCGCG 1641  
 DB 20605 GAGAACTCATGATCTGCTGATCTGTGTGTCGCAACGACCTGTGCGCGGTGTCGCAACGAGC 20664  
 QY 1642 TTCGTCAACGTGCCCCGGCTCTTTCGAGTGAAGACCTTACCGCGCGCTGACACGCTGTG 1701  
 DB 20665 ACCGTGAAGGTACACGCTCTGATGAGCGATGAGTCATATGCGACGCTGCAACGATGTG 20724  
 QY 1702 TCACCATTCGCGGAGACGCGTGTGCGCGCGCGGACCGACCGCGCGCTGCTGATCGCGCGCGC 1761  
 DB 20725 ACCACGCTGCGCGGACGATTCGCTGAGAGAGCTGCGATTGTCGACGTAATGCGAGCGAC 20784  
 QY 1762 TTCGCGCGCGCTTCATGACCGCGCGCGCCCAAGAGCGCACATGAGATCATGACCGC 1821  
 DB 20785 TTCCTGTGTGATGATGATGACCGGGGACCGAAGGAAGCTCCGTAAGAAATCTGGAACGT 20844  
 QY 1822 CTGAGAGAGAGCGCGCGCGCGCTTACTCCGGGCGCTCGATGTGTTGCTCCCTGACGCGC 1881  
 DB 20845 CTGGAAGTTCGCGCGCGCGGAAATCTACTCCGGAATCTCTGGGATATCTCGGCTTCGATCGC 20904  
 QY 1882 GCGCGGACCTCAAGCATCTGATCCGACCATGCTGTGCGCGACCGCGCGCGCGGAGTTT 1941  
 DB 20905 ACCGCGGACCTCAAGCATCTGATCCGACCATGCTGTGCGCGACCGGGTTCGACGATTC 20964  
 QY 1942 GCGCTGCGCGCGCGCGATCTGTCCCTCTCCGACGAGAGAGAGTTCAACGAGACGCTG 2001  
 DB 20965 GGTGCGCGCGCGCGCATTCGTGCGAGCTTCCAGCTGACAGAGATGCGCGCGAGAGAAC 21024  
 QY 2002 GTTAAAGCGCGCGCGCATGTGTCACCGCTTCGACGCGACCGCGCTGTGCGCGG 2051  
 DB 21025 CTCAAAGCGCGCGCGCATGTGCTGTGCGCTGACGCTGTCAACCGCGCGCAGG 21074

## RESULT 10

ACF64464  
 ID ACF64464 standard; DNA; 26309 BP.

ACF64464;

17-OCT-2003 (first entry)

Propionibacterium acnes DNA contig sequence #30.

Acne vulgaris; antiseborrheic; dermatological; antibacterial;

immunostimulant; immune response; vaccine; ds.

XX  
 OS Propionibacterium acnes.  
 XX  
 EN W02003033515-A1.  
 XX  
 PD 24-APR-2003.  
 XX  
 PF 11-OCT-2002; 2002WO-US032727.  
 XX  
 PR 15-OCT-2001; 2001US-00978825.  
 XX  
 PA (CORI-) CORIXA CORP.  
 PI Mitcham JL, Skeiky YAM, Persing DH, Bhatia A, Maisonneuve JL;  
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
 PI Barch B, Valliave-Douglas J;  
 DR WPI; 2003-381789/36.  
 XX  
 PT New Propionibacterium acnes polypeptides and polynucleotides encoding the  
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
 PT or for stimulating an immune response specific for a P. acnes protein.  
 PS  
 PS Claim 1; SEQ ID NO 30; 1481bp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
 CC encoding a Propionibacterium acnes protein. The invention also relates to  
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
 CC immunogenic fragments of P. acnes polypeptides. The invention  
 CC additionally encompasses expression vectors and host cells comprising a  
 CC polynucleotide of the invention; antibodies against polypeptides of the  
 CC invention; fusion proteins comprising a polypeptide of the invention; a  
 CC method for stimulating an immune response specific for a P. acnes  
 CC polypeptide and an isolated T cell population comprising T cells prepared  
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
 CC antigen-presenting cells that express the polypeptide); a method and kit  
 CC for detecting or determining the presence or absence of P. acnes in a  
 CC patient; and a method for inhibiting the development of P. acnes in a  
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
 CC proteins, T cell populations or antigen-presenting cells that express the  
 CC polypeptides are useful for diagnosing, preventing or treating acne  
 CC vulgaris, or for stimulating an immune response specific for a P. acnes  
 CC protein. The polynucleotides can also be used as probes or primers for  
 CC nucleic acid hybridisation. The vaccine composition is useful for the  
 CC stimulation of an immune response against P. acnes, or for treating acne,  
 CC and the kit is useful for performing a diagnostic assay. The present  
 CC sequence represents a P. acnes DNA contig which is specifically claimed  
 CC in the invention. Note: The sequence data for this patent did not form  
 CC part of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 26309 BP; 5546 A; 8173 C; 7885 G; 4659 T; 0 U; 6 Other;  
 Query Match 15.3%; Score 316; DB 8; Length 26309;  
 Best Local Similarity 50.6%; Pired. No. 3,7e-32;  
 Matches 1057; Conservative 0; Mismatches 955; Indels 78; Gaps 9;  
 QY 10 CTTCTGATGAGCAACTGACGCTGTTCAACCAACCTGTTCCAGTACATCGCGAGGCG 69  
 DB 19015 CTGCTGTTGACCAACAGATTCTTCACTTCAACCTGACACCACTATTGGCGGAAGTC 19074  
 QY 70 ACCGCGCAACCCCGCTGCTGTCGCCAAGACGCGACTGTGCG----- 114  
 DB 19075 AGCGGATGAGAGCACTGATGTCGCGCGCGAGAGATGAGGTGCAAGGATCTCCGC 19134  
 QY 115 CGGCTGCGCGTCAAGGACTTTCAGCGGATGCTGTCTCCCGGCGCGCGACCGCGAC 174  
 DB 19135 CGATTGCTGATGAGCAATTCGACCACTGTGTTATGCTGTCAGGACTGCTCCCGCGAC 19194  
 QY 175 CGGGAACGGGACTTTCGGAATTCAGCGCGCGGGGCGATCA-----CGAGACGGCGCTGCGCGCT 231  
 DB 19195 AACGACAAAGACTTTCAGAGCCCGCGACAGGTGATGATGCTTCACAAAGACTCCCGCTG 19254



QY	232	CTCGGCGCTCTGCTCGGCACACGAGGACATGCCAGCTCTTTCGGCGGAAACCGTGGCGCTTC	291
Db	19255	CTCGGGGTGTGCTCTGGGCGACCAAGGCGCTCG---GGTTACGTACCGATGCCAGTTTGACG	19311
QY	292	GCCCCGGAACCCATGCAACGGCGGGGTCTCCGAGGTGGGACACCGGCGAGGACCGCTTC	351
Db	19312	ACCAATTGGGCCCCATCAACGGGATCTGACCGCAATTTATACTTCACAGGCGGTGAATCTTTC	19371
QY	352	CGGGGCGCTCCCTCGCGGTTCAACCGCGGTGCGCTACCACTCCCTGGCGCGCACCGACCTC	411
Db	19372	GCCGGATTACCGCAAGATTTGAGGCAACTGTGGATACCTACTGCTGTTGACCAAGCTTG	19431
QY	412	CCCGACGAGCTGCACCCCTCGCTGAGCGACGACGGGGTGGTCAATGAGGCTTGGCGGAC	471
Db	19432	GGTGAACCGAGTGTCTGCAAACTGCTCGCGCGAGGATGGCGCCACTATGGGCTTTCAGAGTT	19491
QY	472	CGCGAAGAACCGCTGTGGGGGCGTCCGATTCACCCCGGAGTCCATCCGACACCACTTTCGGC	531
Db	19492	GCCGACCGGCGCACTGGGGGGGTGCAAATTCACCCACAGTCACTGACATGACGACGGTGCGA	19551
QY	532	CGGAGATATCATGAGCGCAACTTCCGGAAGCTGACCTCGCCACACCGGGGCAACGGCGCCAC	591
Db	19552	TGTCAATTAAATGAGGAATCTCTGGCGATGGCAGCGCACGACACGACACGACGAGTACGC	19611
QY	592	GGGGCGCACTCCCGTACGAATCC-----ACGTGCGCGCGGTGCAGTGTGCTGGCGG	643
Db	19612	GACAAACGCGAACCGCGCAACTCATGACTCAAGAAAGTGTCCCGGACTACGGGACGGTGG	19671
QY	644	ACGCCGAAGAGTACGCCCGG---CTGCTTGCCTCGGCGAGGGGACCAAGCTTCTGGCTG	699
Db	19672	ACAATTCACCAACCAAGGTGGACATGAGCTCGACGAAAMACCACTTCACGCCCTTC	19731
QY	700	GACAGCAGCTCCGTCTCTGAAAGGGCGCTCGCGCTTCCTCTCTCGGCGACGACCGCGCG	759
Db	19732	GCTGGCGACGGTGACGCTTCTGTGTGACTCGGCGACACGCGTGGCGACACCGGCGGT	19791
QY	760	CCGCTCGCGGAGTACTCACTCAACCGCGTGCACGCGACGCGCTCGTCTCCGTCGCGGACTC	819
Db	19792	TGGAGTGTCAATGGGCAACGGCGTGGGGGGCGGCAATCGCAACTGGTCAAGTTAACAGCTCATC	19851
QY	820	GACGGCACACGACCCCGGACCGCGGCGCCCTTCTTCAACTGAGAGGAGCAGCTCGAA	879
Db	19852	ACAAACACGTTGACGCTCAACGGGCGGAACCCAGCTGTGATGTGCTGCACTTCTTGAA	19911
QY	880	CGCGGAGGGGTCCCGGTGCGCCCGCAACCTGCTCGA-----GTTCAAC	924
Db	19912	CGCGGTCTTGAGATGAGATATCAACGAGGAGACCCGGAAGTATGATATCCATTCA	19971
QY	925	CTCGGCTACGTGCGCTACTTCGAGTACGAGCTGAGGCGGAGACCAACCGCGACCCCGCG	984
Db	19972	GGTGGCTATGTGCGTTTCTGTGGGTATGATGTGCAAGGCGCTGACCTTGGGCGCGAATGG	20031
QY	985	CACCGGTCCCCGCAACCCCGACGCGCGTCTCTTTCGCGGACCGGCGCATGCGCTTGAC	1044
Db	20032	CACCTGACGCAATCGCCGATGCGCATGTGATGCGTCCGGCGACGTGAGTTGGTACGAC	20091
QY	1045	CACACGAGAAGGCTGTGTACTGTGTGGCCCTGACCGCGGGGGCCACGACGACGCGCGC	1104
Db	20092	CACCTACCGGCATCAACGCTCACTCTGACCTTGGACACACGCTATCAACGCGCGCTGCAT	20151
QY	1105	CGCGCTTGGCTGCGGGAGACGCGCGGACCTCAACCGGCTGTGCGCGTCCGCGCGCGCGC	1164
Db	20152	GAGGCGCTGATC-----TATTTGACGCACTGGCTTGCACGCTGGT	20192
QY	1165	GAGCGACCCCGGACATGCTTTCGGGATCCCGAGGCGGCGCGCGCTTTCGCGCCCTTG	1224
Db	20193	GAGGACCAAGTATGACGACGGAAGCGCGGACGCTAACCGTGTCTGTGGAAGGTGTG	20252
QY	1225	GCCGCGCGCGCCACGACAMGAGCGCTTACTTAAGCGCATGACGAGTGTCTTAAAGAG	1284
Db	20253	-----GCGGCTACGCGGAGGGGCTTACAGGACGCTTTCGCCCAATCCACGCGGGG	20304

QY	1285	ATCCGCAACCGCCAGTGTCTGATCGAGATCTGCTCGAACCAATAGTGTACACCGCGCCGACCGAG	1344		
Db	20305	CTAACCGCGCGGTGACTGTGACGAGGCGCTGTCTGATCCGACACGTGAGACTTCCCGGTGCGAC	20364		
QY	1345	GCGACGGCGCCCTCGCGCTCTACTCCGCGCTGCGCGCCATCAGCCCGCTCCGTACGGCGCC	1404		
Db	20365	GTGGAACGGCTGCGCGGTTGTATCCGCGCACTGCGGCGTCGAAACCCAGCCCCCTATGGGGGG	20424		
QY	1405	CTGCTCGAGTTCCCCCGAATCG---TCGGTGTAGAGCGCCTCGCCGAGCGGTTCTCTACG	1461		
Db	20425	TATCTTCGGGTTTACCGACCCGCGCTGTGAGAGGTGTGCTCTTCAATCCCGGAAACGGTTCTCG	20484		
QY	1482	ATCGCGCCCGACCGCGGCGCTCGAGTCCAGCCCATATAAAGGGACCCCGCCCGGGCGGC	1521		
Db	20485	CGAATTACCGCACCGGCATCGCTGAATCCAGCCGATTAAGGGGCAATGGCTGTGTGTAT	20544		
QY	1582	ACCGCGGAGAGAGACGAGCGGCTCCGCGCCGACCTGCGCGCGCGGGAGAAAGACCGGCGC	1581		
Db	20545	GATCCCGGTGAGAGATGCCCGGGGAAATTTGTGATTTTCAACATGACGCGAAAAACCGCGC	20604		
QY	1582	GAGAACTGTATGATCTGTGACACTGTGTCCGACAGACCTCAACAGCGTCTGCGCGATCGGC	1641		
Db	20605	GAGAACTCATGATCGCTGATCTGTGTGCGGACAGACTGTGCGGCGGTGTGCCAACAGAGC	20664		
QY	1642	TCGCTCCACGTGCCCCCGGCTCTTCGAGGTGAGAACCTACGCGCCGTGACACGACTGTGT	1701		
Db	20665	ACGGTTGAGGTACCACTCTGATGTGGCGATCGATCATATGCGACGGTGTGACCAAGATGTGT	20724		
QY	1702	TCGACCAATCCGCGGAGACGGCTGTGCGGCGCGGACAGACACCGCGCGCTGTACGCGCGCGC	1761		
Db	20725	ACCAACGGTGTGCGGAGCATTTGCTGTAGAGACGTGTGATTTGATGACGTACTGTGAGCGACC	20784		
QY	1762	TTCCCGGCGGCTTCATGACCGGCGCGCCCAAGAAACGACACCATGAGATCATCGACCGC	1821		
Db	20785	TTCCCTGGTGGATGTGATGACCGGGGACCGAAGAAACGGTCCGTAAAAATCTCGGACGCT	20844		
QY	1822	CTGAGAGGAAGCCCCCGGGCGCTTACTCCGGGGCGCTGTGATGTTGCCCTCAGCGGC	1881		
Db	20845	CTGAGAGGTGCGCGCGCCCGGGAATCTACTCGGGAATCTGGGATATCTCGGCTTTCATTCGC	20904		
QY	1882	GCGCGCCGACCTCAAGCATGTGTCATCCGCAACATCGTGTGGCCGAGCGGCGAGGCGGATTC	1941		
Db	20905	ACCGCCGACCTCAGCATCTGTGATACGAGCGGTATCCGAACGGGGTGTGCGGTTAAACGGTC	20964		
QY	1942	GCGCTCGGCGGGGCGATCTGTCCCTCTCCGACACAGAGAGAGATTCAACCGAGACCGTG	2001		
Db	20965	GGTCCCGGCGGCGGAATCTGTGACGCTTCCCACTGTGACGAGAGATGTGGCGCAGAAAGAC	21024		
QY	2002	GTAAAGCGCGCGCCATGTGTACCGCCCTTCGACGCGACGCGCGTGGCGGG	2051		
Db	21025	CTCAAAACGGCGGACCGTGTGGCTGTGACGCTGTACGCGGCGCAAG	21074		
RESULT 11					
ID	ACA42265				
AC	ACA42265	standard; DNA; 1362 BP.			
XX	ACA42265;				
XX	19-JUN-2003	(first entry)			
DE	Prokaryotic essential gene #23922.				
XX					
KM	Antisense; ds; prokaryotic essential gene; cell proliferation;				
XX	drug design; gene.				
XX					
OS	Pseudomonas aeruginosa.				
XX					
PN	WO20027183-A2.				
XX					
PD	03-OCT-2002.				
XX					
PF	21-MAR-2002; 2002WO-US009107.				







XX Novel isolated nucleic acid encoding *Pseudomonas aeruginosa* polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
PS Disclosure; SEQ ID NO 13413; 455bp; English.  
XX The invention relates to *Pseudomonas aeruginosa* polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a *P. aeruginosa* nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-*P. aeruginosa* drugs, as templates for recombinant  
CC production of *P. aeruginosa*-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of *P. aeruginosa*-caused  
CC infection, and in detection of *P. aeruginosa* sequences or other sequences  
CC of *Pseudomonas* species using biochip technology. Sequences ABD01397-  
CC ABD1967 represent *P. aeruginosa* polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html  
XX  
SQ Sequence 1101 BP; 158 A; 355 C; 381 G; 207 T; 0 U; 0 Other;  
Query Match 13.0%; Score 268.4; DB 11; Length 1101;  
Best Local Similarity 58.6%; Pred. No. 7,7e-26;  
Matches 486; Conservative 0; Mismatches 341; Indels 3; Gaps 1;  
QY 1227 CCGGCGCGCCGACAGAGACCGCTTACCTCAAGCCGATGACGAGTGGCTTCAAGAGAT 1286  
DB 862 CCGACGAGCATGACGCGATCCAGCTATCGGACGAGCGATCCGCGATCCAGAGCTAAT 803  
QY 1287 CCGCAAGCGGAGTCCGACGATCTGCTCGACCAATGATGATCCGCGCGACCGAGC 1346  
DB 802 CCAAGCGAGGACTGCTACCAAGGAACTATAGCAGAGCGTTCCAGAGCCGCTTGAGCGG 743  
QY 1347 GACGGCCCTCGCTTACTCCGCGTCCGCGCATGAGCCCGTCCGATGAGCGCCCT 1406  
DB 742 CTGCGCGTGGCGGCTTATGCGCGCTGCGGAGCGCTGCGACGCGCTTCTCCGCTA 683  
QY 1407 GCTGAGTTCCCGAATGCTGCTGAGCGCTGCGCGAGCGGTTCTTCAAGATCGG 1466  
DB 682 CTGCGACTGAGCGAGCGCGATCTCAGCTGTGCGCGAGGCTTCTCAAGCT--- 626  
QY 1467 CGCGGAGCGGCGGCTGAGTCCAAAGCCATCAAGAGGACCCGCGCGCGGCGACCGC 1526  
DB 625 CGACAAAGGCGAAGGTGAAACCGGCGGATCAAGGGGACCCGCGCGGCAAGACCC 566  
QY 1527 GGAAGAGAGAGAGCGGCTCGCGCGCGACTGCGCGCGCGGAGAGAGACCGGCGAGAA 1586  
DB 565 CGAGAGAGAGATGCGCGCTGCGCGCGCTGCTGCTGCGCGCGCGAGAGACCGCGGAGAA 506  
QY 1587 CTTGATGATGCTGACCTGCTGCGCAAGCCTCAACAGCGTCTGCGCGATGCGCTCGT 1646  
DB 505 CTTGATGATGCTGACCTGCTGCGCAAGCAGATGAGAGTTGCCAACCTGGACGCGT 446  
QY 1647 CCACGTGCGCGCGCTTCTTGAAGTGGAGACCTACGCGCGCGTGCACCAAGCTGTTGAC 1706  
DB 445 ACGGATACCGAGAGCTGTCGCCCTCGGAAACCTATCCCAAGTGTGATCACTGTGAGCAG 386  
QY 1707 CATCCGAGGAGCGGCTGCGCGCGCGGACACAGACCGCGCGCTGATGCGCGCGCTTCCC 1766  
DB 385 GGTACCGGAGAACTGCGCGCGCGGCAAGAGCGCTTGTGACCTGTTGAAGAGCGCTTCCC 326  
QY 1767 CCGCGGCTTCATGACCGGCGCGCGCCCAAGAGCGCAGCATGAGATCATTCAGCGCTTGA 1826  
DB 325 CCGGCGGCTTCATGACCGGCGCGCGGCAAGATTGCGGCGCATGACATGACAGCACTGGA 266  
QY 1827 GGAAGGCGCGCGGCGCTTACTTCGCGGCGCTGAGATGTTGCGCTTCAAGCGCGCGC 1886  
DB 265 ACCGAGCGGAGCGGCGATCTTACTGCGGCGAGCGCTTCTACTCTGACGCTGCGCGCGAGAT 206

QY 1887 CGAAGCTGATGCTATCCGACCATGCTGTCGCGCGAGCGCGAGCTTGGCGT 1946  
DB 205 GGAAGCTGATGCTATCCGACCATGCTGTCGCGCGAGCGCGAGCTTGGCGG 146  
QY 1947 CCGCGGCGGATGCTGCTTCCGACCAAGAGAGAGTTCACCGAGACCGTGTAA 2006  
DB 145 CCGCGGCGGATGCTGCTGCGGCTGCGGAGAGAGAGTTCACCGAGACCGTGTAA 86  
QY 2007 GCGCGCGGATGCTGCTGCGGCTGCGGAGAGAGAGTTCACCGAGACCGTGTAA 2056  
DB 85 GGTGCGGCTGCTGCGGAGAGAGAGTTCACCGAGAGAGTTCACCGAGACCGTGTAA 36  
RESULT 13  
ABD15129  
ID ABD15129 standard; DNA; 1431 BP.  
XX  
AC ABD15129;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE *Pseudomonas aeruginosa* polynucleotide #13733.  
XX  
KW Bacterial infection; gene; ds; *Pseudomonas aeruginosa* infection;  
KW antibacterial.  
XX  
OS *Pseudomonas aeruginosa*.  
XX  
PN US651795-B1.  
XX  
PD 22-APR-2003.  
XX  
PF 18-FEB-1999; 99US-00252991.  
XX  
PR 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
DR WPI; 2003-615309/58.  
DR P-PSDB; AB081558.  
XX  
PT Novel isolated nucleic acid encoding *Pseudomonas aeruginosa* polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
PS Disclosure; SEQ ID NO 13733; 455bp; English.  
XX The invention relates to *Pseudomonas aeruginosa* polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a *P. aeruginosa* nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-*P. aeruginosa* drugs, as templates for recombinant  
CC production of *P. aeruginosa*-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of *P. aeruginosa*-caused  
CC infection, and in detection of *P. aeruginosa* sequences or other sequences  
CC of *Pseudomonas* species using biochip technology. Sequences ABD01397-  
CC ABD1967 represent *P. aeruginosa* polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html  
XX  
SQ Sequence 1431 BP; 262 A; 503 C; 441 G; 225 T; 0 U; 0 Other;  
Query Match 13.0%; Score 268.4; DB 11; Length 1431;  
Best Local Similarity 58.6%; Pred. No. 7.3e-26;  
Matches 486; Conservative 0; Mismatches 341; Indels 3; Gaps 1;



OY	1227	CCGCGCGCGGCCA	CGACAAGGAGCGCTTACCTTAAAGGCGATTCGACGAGTGCCTTCAAGAGAT	1288
Db	597	CCGACGGAGCAT	CACCGCATCCGACTTATGCGACGCGCATTCGCGCATTCAGAGACTTACAT	656
OY	1287	CCGCAACGCGCGAT	CGTACGAGATCTGCTTGAACCAATGATCAACCGCGCGACCGAAGGC	1348
Db	657	CCAGGACAGCGCAT	CTGCTACAGGGTGAATATATAGCAGGCGTTCCAGGCGCGTTTCAGAGG	716
OY	1347	GACGGCCCTTG	CCGCTTACTCCGGCTGCGGCCATCAAGCCCGTCCGTACAGCGCCTT	1406
Db	717	CTCGCGGTGCG	CGCGCTTATCGGCGCTTGGCGACAGGCTGCGGACGCGGTTCTCCGCTA	776
OY	1407	GCTGAGTTCC	CGAATGTGGTGTGAGAGCGCTTGGCCGAGCGGTTCTTCAAGATCGG	1466
Db	777	CTCGGACTGG	CGCGCGCATCTTCAAGCTTGCGCGAGCGCTTCTTCAAGCT--	833
OY	1467	CGCCGACCGCG	CGTCTGAGTCCAGCCCATCAAGGGAGACCGGCCCGCGGCGGACCGC	1528
Db	834	CGACAGGGGCA	AGGTGAAACCGGCGCATCAAGGGGACCGCGCGCGCGGAGAACCC	893
OY	1527	GAGAGAGACG	AGCGGCTTCCGCGCCGACTTGCGCGCGCGGAGAAAGACCGGCGGAGA	1586
Db	894	CGAGGAGGACA	TGGGGCTGCGCGGCTTCCGCTGTGCGCGCCCAAGGACCGCGCGGAAA	953
OY	1587	CCTGATGATCG	CGACTGTGGTCCGCAAGACCTCAAGCGGTCTGGCGGATCTGGCTCCG	1646
Db	954	CTGTATGATCT	GCTGACTCTGCTGCGGACAGACATTCGACGCACTTGCACACTGGAGCGT	1013
OY	1647	CCAGCTGCCCC	CGGCTCTTTCGAGGTGAGACCTACGCGCGGTGACACAGTGTGTGAC	1706
Db	1014	ACGGGTACCG	AGGTCTTTCGCCCTGAAAGCTATCCAACTGATCACTGTGTGAGCAG	1073
OY	1707	CATCCGGGGA	CGGCTGCGGCCGACACAGCAGCGCGCTTGCGTACGCGCGCGCTTCC	1766
Db	1074	CGTCACTCGG	CGAATCGGCGCGGCGGACAGCGCTTCACTGTGGAAGGCAAGCTTCC	1133
OY	1767	CGGCGGCTCA	TGATCCGCGCGGCCCAAGAGGACATATGAGATCATCGACCGCTGGA	1826
Db	1134	CGGCGGCTCG	ATCATCCGCGCGCGGACGAAATTTGCGGCATTCGACATCATTCGACGA	1193
OY	1827	GGAAGGCCCC	CGGCGGCTTACTCTCGGCGCGCTCGATGTGTGCGCTTCAAGCGCGCGC	1886
Db	1194	ACCGAGCGCA	GCGCGCATCTACTGTGGCAGGCTGTCTTCACTGACGTGCGCGCGAGAT	1253
OY	1887	CGACTCAGCA	TCTGTCATCCGACCATGTGTCTGGCGACGCGCGAGCGGATTCGCGCT	1946
Db	1254	GGAACGCTCG	ATTCGCTATCCGACCTGCTGCTGATCAGAAACGCGCAGGTCAATTCCTGGG	1313
OY	1947	CGGCGGGCGAT	TGCTGCTTCTTCGACGAGAGAGAGATTCAACCGAGACGCTGTAA	2006
Db	1314	CGGCGGCGGCA	TCTGTCGCTGACTGCGACTTGGAGAGACGATGACAGGAAACCTTGAGACA	1373
OY	2007	GCGCCGCGCAT	GTGTGTCACCGCCCTTCGACGCGACGCGCGTGGCGGAGCGCC	2056
Db	1374	GGTCTGGGTG	CTGTGTAACCTTGGAAAGGATGGCCGCGGACAGGTCCTC	1423
RESULT 14				
ADBB06077				
ID	ADBB06077	standard; DNA; 2046 BP.		
XX	ADB06077;			
DT	20-NOV-2003	(first entry)		
XX				
DE		Alloicoccus oclitis antigenic protein encoding DNA SEQ ID NO:17.		
XX				
KW		Alloicoccus oclitis; antigenic protein; immunogenic; immunisation; gene therapy; Gram-positive bacterium; infection; gene; de.		
XX				
XX		Alloicoccus oclitis.		
XX				

WO2003048304-A2.

12-JUN-2003.

25-NOV-2002; 2002MO-US036123.

29-NOV-2001; 2001US-033777P.  
18-NOV-2002; 2002US-0426742P.

(AMHP) WYETH HOLDINGS CORP.

Fletcher LD, Mcmichael JC, Russell DP, Zagursky RJ;  
WPI; 2003-505284/47.  
P-PsDB; ADB06078.

New Alloiococcus otitidis polynucleotides and polypeptides, useful for treating and diagnosing diseases, drug screening assays and monitoring of effects during drug clinical trials.

Claim 7; SEQ ID NO 17; 1019pp; English.

The present invention describes an isolated polynucleotide (I) of Alloiococcus otitidis genomic DNA, which encodes an antigenic protein. Alloiococcus otitidis is a Gram-positive bacterium. Also described: (1) an isolated polypeptide that is encoded by the polynucleotide (1); (2) an expression vector comprising the novel isolated polynucleotide (1), its complement, degenerate variant or fragment; (3) a genetically engineered host cell, transfected, transformed or infected with the vector of (2); (4) an antibody specific for the polypeptide of (1); (5) an immunogenic composition comprising the polypeptide, its complement, biological equivalent or fragment, or the polynucleotide that is comprised in the expression vector; (6) a pharmaceutical composition comprising the polypeptide of (1) and a carrier; (7) a protein chip comprising an array of the polypeptides of (1), their biological equivalent or fragment; (8) immunising against Alloiococcus otitidis by administering to a host the immunogenic composition; (9) detecting and/or identifying Alloiococcus otitidis in the biological sample; (10) a kit comprising a container containing the novel polynucleotide, its degenerate variant or fragment, or the antibody of (4); and (11) producing a polypeptide by culturing the genetically engineered host cell under conditions suitable to produce the polypeptide from the culture. (1) can be used in gene therapy. The polynucleotides, polypeptides, antibodies and compositions of the present invention can be used for treating and diagnosing diseases, drug screening assays and monitoring of effects during drug clinical trials. The polynucleotides are useful for expressing and detecting Alloiococcus otitidis. The present sequence encodes an Alloiococcus otitidis antigen protein from the present invention.

Sequence 2046 BP; 597 A; 470 C; 493 G; 486 T; 0 U; 0 Other;

Query Match 13.0%; Score 267.2; DB 9; Length 2046;  
Best Local Similarity 48.8%; Pred.No. 9.66-26;  
Matches 999; Conservative 0; Mismatches 963; Indels 84; Gaps 7;

1 ATGGCGACGCTTCGATCGACAACTAGACTGTTCAACCACAACCTGTCCAGTACATC 60  
Db | |||||  
22 ATGAAGTCATCATTAATCGAATAACATGATTCTTATACCTTAACAATTATACAGCTTAT 81  
Oy | |||||  
Oy | |||||  
Db | |||||  
121 CCCGTGCA--GCACTTGACGCGATGTCGTGTCGCCGAGCGGCAGCTGTCGGCGCTG 120  
Oy | |||||  
Db | |||||  
142 CTATTGATCTCGACTTGTAAATAGTATTAATTTCACCGGGGCCGGTAGTCCGAGACGG 201  
Oy | |||||  
Db | |||||  
178 GAACGGGATTTGGAAATACAGCCGCCGGGCGCATCACGACAGCGCCCTGCTCTCGGC 237  
Oy | |||||  
Db | |||||  
202 GACAAGGACTTCGCTGTGTCGGCAGATGAATGAAAAGCTTGACAGAAGCCATATCTTTGGC 261  
Oy | |||||  
238 GTCTGCTCGGCGCCACGAGGATATGCGACAGCTTTCCGGCGGAACCGTTCGAGCTCGCCCG 297



Db 262 ATTGCTGGGCCACGAGGCATTTATTATTAAAGCGTGCACCTTAGTGCGGCGCTGAT 321  
Qy 298 GAACCCATGACGCGCGGGTCTCCGAGGTGCGGCACACCGCGAGAGACGTCTTCGGGGC 357  
Db 322 ATCCCATGATGATGGCGGAGAGTTCGTCTACCATATATGGAAAAATCTTTGACGCG 381  
Qy 358 CTCCCTCGCGTTTACCGCGTGCCTACCTCCCTGG--CCGCACCGACCTTCC 414  
Db 382 ATAGACCAAGGCTTTGAAGTCAACCGCTACCACTCTTGTGTGAAGACAAGAAATTA 441  
Qy 415 GAGAGCTGACACCCCTCGCTGAGAGAGACGAGGGTCTGCTATGGGCTCGCGCACCGC 474  
Db 442 GAGATATTTAAATTCATGTCGCAAAAACGATGATGAATGTCTATGCGCTGTCCCAAG 501  
Qy 475 GAGAAAGCGCTGTGGGCGTGCAGTTTCACCCGAGTCCATCGACGCACTTCGCGCG 534  
Db 502 ACCAAGCCATTCATGAGGATCCATTTCACCCGAAATCATTTGGCACCCAGTATGGGAA 561  
Qy 535 GAGATCATGGCCAACTTCCGCGACCTCGCCCTGCCCAACCGGGGACGCGCGCACGG 594  
Db 562 AAATGATGAAAACTT-----TATGGCTTTGTCCAGGAC 597  
Qy 595 GCGACCTCCCGTACGACATCCAGTGCAGCGGTGACGTGCTCGCGCGCGACCGCGAAG 654  
Db 598 TACTATAACCAATCCAGCTTTTATTATGAAAAAGTTCGGAACTAGGGACACGCAAGAC 657  
Qy 655 GTACGCGCGGCTGCTGCGCGGAGGAGCACCACTTCTGCTGACAGACAGCTCCGT 714  
Db 658 CTCTACATCATTTGGCGCAAAATTTGATGACCAAGTACTTTGGCTCGACTCGACAGAGT 717  
Qy 715 CTGAGAGCGCTTCGCGCTTCTCTTCTCTCGGCGACGACCGCGCGCTCGCGGATAC 774  
Db 718 GAGCTGGCTTGTACCGTTTCCATCTTTGGCATGTGACCAACCAAGCGGGCCACAC 777  
Qy 775 CTACCTACCGCGTGCAGCGGCGTGTCTCGTCCGCGCGCTCCGACGCGCACACGAC 834  
Db 778 TTTAAATATGATGTGACGCAAAAAGATGAAAAAAGATGACGTCGCGACGATT 837  
Qy 835 CGGACGCGCGCGCTTCTTCAACTACCTGAGGAGCAGCTCGAACGCGACGCGGTCCC 894  
Db 838 GAAA---GCTTCCAGATGACATCTTTCTTAACTTAAAGCCAAACGACCAAAATGGCT 894  
Qy 895 GTGCGCCCGCACTGCGCTTGAAGTTCACTCGGCTACGCTACCTCGGCTACG 954  
Db 895 TATGAGCGCGCTTGCCTTGTGACTTTCAGCTAGGCTATATGCTATATCGCTATGAG 954  
Qy 955 CTGAGGCGGAGAC---CACCGCGACCGCGGACCGGTCGCCGACCCCGACCGCGCG 1011  
Db 955 GTCAAAAAGATATGATTCAAGTTCACACAGACCAAGTACCTATTCGAGATGCTTAC 1014  
Qy 1012 TTCTCTTTCGCGACGCGGCATGCGCTCGACACACGAGAGGCTGCTACTGCTG 1071  
Db 1015 TTTACTATTGTGACCGGGCTTGTGTATGACACAGAGAGGACTTTTACTTCTC 1074  
Qy 1072 GCGCTGACCGCGGGGCAACGACGACGCGCGCGCTGCGTGGCGGAGACGCGCGAG 1131  
Db 1075 TCTTACCA-----GGAATGAC 1089  
Qy 1132 ACCCTCACCGGCTTGGCGGTCCGCGCGCGCGGACCGACCCCGGCATGATCTTCGG 1191  
Db 1090 CAGATTGATTCAAGGGTCAAAAGACAGGCTCAAGGACAGTCAAAAGTTACAGCAAA 1149  
Qy 1192 ATCCCGAGGCGCGCGCTTGGCGCTTGGCGCGCGCGCGCGCGCACGACAGAGCGCC 1251  
Db 1150 GGAACCAAGGACGAAAGACTTTCACGACTGGCTTTCG---TCAAGACAAAGCAGCC 1206  
Qy 1252 TACCTCAAGCGCATGACGAGTGTCTCAAGAGATCCGCAACGCGGAGTGTGAGATC 1311  
Db 1207 TATATCAAGGATTTAAACCATCATCAGACCTGATCAAGGCTGGGAAAGTTATGAAGT 1266  
Qy 1312 TGCTGTACCAATGTGTACCGCGCGACGAGGCGACGCGCTCGCGCTTACTCGCG 1371  
Db 1267 TGCTGTACCAACCGCGCTGATATTGAGGATCAGATTGATGCGCTCTTATTACAAACAC 1326

Qy 1372 CTGCGGCCATGAGCCCGCTCCGTAACGCGCGCTGCTGAGTTCCCGAACTGTGCTG 1431  
Db 1327 TTGCGCCAGAAAGTCCCGGTCAGTACTGTGCTTTTTCCTTTAGATGATTTTCAAGTT 1386  
Qy 1432 CTGAGGCTCTCGCCGAGGCTTCTCAGATTCGCGCGGACGCGGCGCTCGATCTAAG 1491  
Db 1387 CTTCCTCTTTCATGAGAGGTTTATTATATGTGACAAAGACCGGCTTGTGACACTAAG 1446  
Qy 1492 CCCATCAAGGGAGCCCGCGCGCGGCGCACCGCGGAGAGAGACGAGCGGCTCGCGCG 1551  
Db 1447 CCCATCAAGGGAGCGGTGACGCGGGGACAGATGACCAAGAAACCAACTTGTATGAA 1506  
Qy 1552 GACCTGCGCGCGGAGAAAGACCGGCGGAAACTCTGATGATGCTGACCTGTGCGC 1611  
Db 1507 GGCCTCCGAGTGAAGAAACCAAGGCGAGAAACTGATGATGTGACCTTTCGCGC 1566  
Qy 1612 AACGACTCAACAGCTCTGCGCATGCGCTCCGTCCACAGTCCCGCGCTTTCAGAGT 1671  
Db 1567 AATGATCTGGGCGCTTTTGTGAAATCGGGTGGTGAAGTACCCAACTGATGATGTA 1626  
Qy 1672 GAGACCTACGCGCGGTGACACCAAGCTGTGTGACCATTCGGGGAACGCTGCGCGCGC 1731  
Db 1627 GAGACTTACTTCACTCCCTCCACAGCTGTGTTACACGCTTTCAGCGCGGTCAAAAGAC 1686  
Qy 1732 ACCAGACCGCGCGCTGACGCGCGCTTCCCGCGGCTTCATGACCGCGCGCGC 1791  
Db 1687 CTAGATGTGTTGAAGTGTATGAAACACTTCCCGGGGCGGTTCATGCTGGGCGCT 1746  
Qy 1792 AAGAACGCAACATGAGATCATCGACCGCTGAGAAAGGCCCGCGGCGCTTACTC 1851  
Db 1747 AAAAAAAGAACCTTGAAATTTATGATGACTTGAAGAGCTTCCAGAGGAGATCTATTCT 1806  
Qy 1852 GGGGCGCTCGAATGTTGCGCCTCAGCGGCGCGCGCACTGAGATGCTCATCCGAC 1911  
Db 1807 GGGACTATTGGCTTTTATGACCAACAATTCACATATGACCTTCAACATGCTATCCGAGC 1866  
Qy 1912 ATGCTGCTGCGGACGCGCGGAGGATTCGCGTTCGCGGCGGAGATCGTGTCTTCC 1971  
Db 1867 AGTGTGTGAAGCAGACAAAGGCAAGCATCGCGCTGGGGGTGCGATTGTCAATGCTATCT 1926  
Qy 1972 GACCAAGAGAGAGATTCACCGAGACCGTGTAAAGCCCGCGGCTATCGCGCTC 2031  
Db 1927 GATCTCGAAGAAAGATTGATGAGTTGTTTAAAGCTTAAGGGGCGCTTGTCCGCTTG 1986  
Qy 2032 GACGCG 2037  
Db 1987 CAAAGC 1992

RESULT 15  
ADBI2064.00/c  
WP Sequence split into 18 fragments Locus ADBI2064 Accession Adbi2064  
WP Fragment Name Begin End  
WP ADBI2064\_00 1 11000  
WP ADBI2064\_01 100001 210000  
WP ADBI2064\_02 200001 310000  
WP ADBI2064\_03 300001 410000  
WP ADBI2064\_04 400001 510000  
WP ADBI2064\_05 500001 610000  
WP ADBI2064\_06 600001 710000  
WP ADBI2064\_07 700001 810000  
WP ADBI2064\_08 800001 910000  
WP ADBI2064\_09 900001 1010000  
WP ADBI2064\_10 1000001 1110000  
WP ADBI2064\_11 1100001 1210000  
WP ADBI2064\_12 1200001 1310000  
WP ADBI2064\_13 1300001 1410000  
WP ADBI2064\_14 1400001 1510000  
WP ADBI2064\_15 1500001 1610000  
WP ADBI2064\_16 1600001 1710000  
WP ADBI2064\_17 1700001 1754382  
ID ADBI2064 standard; DNA; 1754382 BP.



XX ADB12064;  
AC 20-NOV-2003 (first entry)  
XX 20-NOV-2003 (first entry)  
DE Allolococcus otilis entire genome sequence SEQ ID NO:6651.  
XX Allolococcus otilidis; antigenic protein; immunogenic; immunisation;  
KM gene therapy; Gram-positive bacterium; infection; gene; ds.  
XX Allolococcus otilidis.  
OS Allolococcus otilidis.  
PN WO2003048304-A2.  
XX 12-JUN-2003.  
PD 25-NOV-2002; 2002WO-US036123.  
PF 25-NOV-2002; 2002WO-US036123.  
XX 29-NOV-2001; 2001US-0333777P.  
PR 18-NOV-2002; 2002US-0426742P.  
XX (AMHP) WYETH HOLDINGS CORP.  
PA Fletcher LD, McMichael JC, Russell DP, Zagursky RU;  
XX WPI; 2003-505284/47.  
XX New Allolococcus otilidis polynucleotides and polypeptides, useful for  
PT treating and diagnosing diseases, drug screening assays and monitoring of  
PT effects during drug clinical trials.  
XX Example 3; SEQ ID NO 6651; 1019p; English.  
XX The present invention describes an isolated polynucleotide (I) of  
CC Allolococcus otilidis genomic DNA, which encodes an antigenic protein.  
CC Allolococcus otilidis is a Gram-positive bacterium. Also described: (1)  
CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an  
CC expression vector comprising the novel isolated polynucleotide (I), its  
CC complement, degenerate variant or fragment; (3) a genetically engineered  
CC host cell, transfected, transformed or infected with the vector of (2);  
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic  
CC composition comprising the polypeptide, its complement, biological  
CC equivalent or fragment, or the polynucleotide that is comprised in the  
CC expression vector; (6) a pharmaceutical composition comprising the  
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array  
CC of the polypeptides of (1), their biological equivalent or fragment; (8)  
CC immunising against Allolococcus otilidis by administering to a host the  
CC immunogenic composition; (9) detecting and/or identifying Allolococcus  
CC otilidis in the biological sample; (10) a kit comprising a container  
CC containing the novel polynucleotide, its degenerate variant or fragment,  
CC or the antibody of (4); and (11) producing a polypeptide by culturing the  
CC genetically engineered host cell under conditions suitable to produce the  
CC polypeptide from the culture. (1) can be used in gene therapy. The  
CC polynucleotides, polypeptides, antibodies and compositions of the present  
CC invention can be used for treating and diagnosing diseases, drug  
CC screening assays and monitoring of effects during drug clinical trials.  
CC The polynucleotides are useful for expressing and detecting Allolococcus  
CC otilidis. The present sequence represents the entire genome of  
CC Allolococcus otilidis, which is given in the exemplification of the  
CC present invention.  
XX  
XX  
SQ Sequence 1754382 BP; 484756A; 391265C; 387369G; 490992T; 0U; 00Other;

Db 7747 GGGAGGTCACAGGGAAGAGCCCATGTGTATTAATAAGACGAGATGACCTACCAAGAA 7688  
Qy 121 CCCGTGCA---GAGCTTCACGCGATGCTGTGTCCTCCCGGGCCCCGAGCCCCGACCGG 177  
Db 7687 CTATTGATCTTCACCTTGTATTAATGTCATATTTCACCGGGGCCCGGTAGTCCGACCGG 7628  
Qy 178 GAAAGGACTTCGGAATACAGCCGCGGGGAGATACCGACGAGCGGGCCGCTCCGCTCGG 237  
Db 7627 GACAAAGACTTCGAGCTTGTCGCGGCAATTAATGAAAGCTTACAAAGCTTATCTTGGC 7568  
Qy 238 GTCGCTCGGCGCAGGAGGATGCGCCAGCTTCGCGCGGAAACGTCGCGCTCGCCCG 297  
Db 7567 ATTGCTCGGCGCAGGAGGATTTATTAATGAGGTCGAGCACTTAGTGGGGGCTGAT 7508  
Qy 298 GAAAGGACTTCGAGCGGCGGCTTCGAGGTGCGGCAACCGGCGAGAGCTTTCGCGGCG 357  
Db 7507 ATCCCATGATGATGTCGCGGAGAGTGTGCTACATTAATGGAATAATATCTTGAACGCG 7448  
Qy 358 CTCCTCCGCGCTTCACCGCGCGGCTGACCTCCCTGG---CCGCGACCGACCTCCCG 414  
Db 7447 ATGACCAAGGCTTTGAAAGTCAACCGCTACGACTCAATGTTTGTGAAGCAAGAAATTA 7388  
Qy 415 GACAGCTCGAACCCTTCGCTGAGAGGAGAGCGGGGTGTCATGAGGCTCGCGGACCGG 474  
Db 7387 GAGATATTCATGATGATGATGCAAAACGATGATGAAATGTCATGGCCCTGTCACCAAG 7328  
Qy 475 GAGAGCGCTGTGGGCGCTCCAGTTCACGCGAGTTCATGCGGACGAGCTTGGCGCG 534  
Db 7327 ACCAAGCCCATGATGAGGCTGCAATTCACCCGATTCATGCGGAGTATGGGAA 7268  
Qy 535 GAGATCATGCGCAACTTCGCGGACCTGCGCCACACCGGCGGACCGGCGGCGG 594  
Db 7267 AAATATGAGAAACTT-----TATGCTTTGTGTCAGAGAC 7232  
Qy 595 GCCGATCCCGGATGCAACTTCACGCGCGCGTGCAGTGTGTCGCGGACCGGACGAGAG 654  
Db 7231 TACTATACCAATTCAGCTTTATTAATGAAAGTTCGCGGAACTAGGAGACCAAGAC 7172  
Qy 655 GTAGCGCGGCTGCTGCTGCGCGGAGGAGACCACTTGTGTCAGACAGCTTCGCT 714  
Db 7171 CTCTAGATCAATTTGCGCAAAATGATGACCAAGTACTTGTGCTGACTTCGCAAGAGTG 7112  
Qy 715 CTCGAAGGCGCTCGGCTTCCTTCCTGCGGAGACGAGCGCGCGCTGCGCGAGTAC 774  
Db 7111 GAGCTGCTGTCACGCTTTTCATCTTTGATGCTGAGCAAAAGGAGGCGCCACACC 7052  
Qy 775 CTCACCTACCGGTCGCGCAGCGCGCTGCTCGGCGCTGCGAGCGGACACGACG 834  
Db 7051 TTAATATGATGATGTCACCAAAAGAGTGAAGAAACAGATAGCCAGTCCGACGAT 6992  
Qy 835 CGAGCGGCGCGCTTCTTCAACTGAGAGACGAGCTGMAACGCGAGCGGCTCCC 894  
Db 6991 GAAA---GCTTCACAGATGACATCTTTCTTAAAGCCAACGACCAAAATAGGCT 6935  
Qy 895 GTGCGCGCGGAGCTGCTTCAGATTCACCTGCTAGTGTGCTGCTGCTGCTGCTGCTG 954  
Db 6934 TATGAGCGGCTGCTGCTTTCATCTTGAAGTATGATGCTATATGCTGCTATGAG 6875  
Qy 955 CTGAAGGCGGAGAC---CACCGGACCGCGGCGAGCTGCTGCTGCTGCTGCTGCTGCTG 1011  
Db 6874 GTCAAAAGATACAGTTAATGTCACCAAAAGACACAGCTACCTATTCGATGCTTAC 6815  
Qy 1012 TTCCTTTCGCGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1071  
Db 6814 TTTACTTATTTGAGACGCGGCTTGTGCTATGACACAGAGAGAGAGACCTTATCTTCTG 6755  
Qy 1072 GCGCTGACCGCGGCGGCGGACGAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 1131  
Db 6754 TCTTACCA-----GGATGAC 6740  
Qy 1132 ACCCTTACCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1191  
Db 6739 CAGGATGATTCAGAGCGGTCAAGACAGCTCAAGCAGGAGTCAAGATTGACAGAA 6680











Db 311 GGGTGTGCTGGGGCAACAGACCCTGTGTCTGTCGGGCGCCGCGCTGTCTCCAGCAC 370  
Qy 296 CGGAACCCATGACAGCGCGGGGATCTCCGAGGTGCGGACACACGGGAGAGATCTTCCGGG 355  
Db 371 CGGAACCCCTTTACAGCGCGGACAGGACATCCGCCAGACGAGGAGAGGCTGTTCGGA 430  
Qy 356 GCGTCCCTGCGCGTTTACCGCGGTGCGTACCATCTCCGAGCGGACCACTCCCG 415  
Db 431 ACATCCCTCCCGCTGACCGTGGTCCGTTACCACTGCTGACCGCTCCGCAACTGCGCG 490  
Qy 416 ACCGAGCTGAAACCCCTGCTGCGGAGCGACAGCGGGGTGCTCATGAGGCTGCGGACCGG 475  
Db 491 CCGACCTGCGCGCCACCGCCACACCGCCGAGCGGAGCTGATGCGCTGCCCGCCAGCC 550  
Qy 476 AGAAGCGCTGTGGGGGCTGCAAGTTCCAGCCGAGTCCATCGGAGGACCTTTCGGCGGG 535  
Db 551 ACCTGCGCCGCTTCCGCTGCTGAGTTCCACCCCGAATGATGACGAGCAACGGCCACC 610  
Qy 536 AGATCATGACCAACTTCCGCGACT----- 560  
Db 611 GGAATGCTCGCAACTTCCGCGACTGTCCCTGCGCGCGCGGCGCACCGCCCGCGACA 670  
Qy 561 -----CGCCCTCGCCACACCGGAGACCGGCGCCACCGGGCCGACTCC- 604  
Db 671 CCGAACCGATACCCCGACCCCGACCCGCCCCCGCCCGCCCGACCGGACCGGCG 730  
Qy 605 -----CGTACGAATCCACCGTGGCGCGGCTGACGTGCTGCGCGGAG 646  
Db 731 CGTCCGCGCGGATGAGGAGGAGTACCGGCTGATGTGCGGAGGTGCGCTGCTGCGGAG 790  
Qy 647 CCGAAGAGATACCGCGCGGCTGCTGCGCGGAGGAGGACACAGTTTGTGCTGAGACGA 706  
Db 791 CCGACGCGCGGATTCACCGCCCTGTTGCGCGACCGCGCGCGGCTTGTGCTGAGACGA 850  
Qy 707 GCTCCGCTCTGAGAGGCGCTGCGGCTTCTTCTTCTGCGGAGACCGGCGCGCTG 766  
Db 851 GCGCGCTGAGCGCGGCTGCGCGCTTCACTTCTTCTGCGGCGCGCGCGCGCTG 910  
Qy 767 CCGAGTACTCTACCTACCGGCTGCGCGGAGCGGCTGCTGCTGCGCGGCTGCGGAGGCA 826  
Db 911 GCGAAGAGATCTACTTACGATGTGCGGACCGGCGCGGTGCGGCTGAGAGAGGTTCAAGG 970  
Qy 827 CCAAGACCGGACCGCGCGGCTTCTTCACTTCACTGAGAGAGCAGCTGAAACCGCGAC 886  
Db 971 GCGAGACCGCGCGGCGGACCTTTCGACACCTGGAACAGAACTGG---CGGCC 1027  
Qy 887 GGGTCCCGTGGCGCGGAACTGCGCTTTCGAGTTCAACTTCGCTGATGCTGCTGCTG 946  
Db 1028 GCGCGCTGCGCGGACCGGCTGCTTTCGAGTTCAACTTCGCTGATGCTGCTGCTG 1087  
Qy 947 GCTACGAGCTGAGAGGCGGAGACCAACCGGAGACCGCGGAGCGGATCCCGGACCGCGAG 1006  
Db 1088 GCTACGAGACCAAGGCGGAGACGCGGCGGAGAGAGCGCCACCGGCGGAACTGCGGAG 1147  
Qy 1007 CCGCGTTCTTTCGCGGACCGGCGCATGCGCTTCGACCAACAGAAAGCTGTGCTACC 1066  
Db 1148 GCGCGCTTTCATGTTTCGCGGACCGGAGTGTCTGCGCTTCGACCAACAGGCGGCGCTG 1207  
Qy 1067 TGTGCGCTTTCGACCGCGGAGGCGACGAGAGCGGCGCGCTGCTGCTGCGGAGAGCG 1126  
Db 1208 TCTGCGGACCTGAGAGAGACCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1252  
Qy 1127 CCGAGACCTTTCACCGGCTGCGGCTGCGGCGCGCGGCGGAGCGGAGCGGCGGCGGCTG 1186  
Db 1253 CCGAAGCTTGGCTTACCGAGCGCGCGCGGAGCTTTCGCGGAGCGGCGGCGGCGGCGGCT 1312  
Qy 1187 TCGGAGTCCCGAGGCGGCGGCTTTCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGAG 1246  
Db 1313 TCAACCTGCTGCGCGAGAGCAACTGCGCGCGCTGAGAGCTGCTACGCGGCGGAGCGG 1372  
Qy 1247 ACCGCTACTTCAGAGCGGATGAGAGTGTCTTCAAGAGATTCGCGAGCGGAGTGTGAG 1306  
Db 1373 CCGGCTAACCGGAACTGTGAGAGATGCGCGCGCTGATACCGAGCGGAGAGCTGAG 1432

Qy 1307 AGATCTGCTGACCAATATGTATCCGCGCGACCGAGGCGAGCGGCGCTGCGCTACT 1366  
Db 1433 AGGTGTGCTGACGAATATGCTTCGGGTGCGCGGCGGATTCGACCGCTTCAACCGCTTAC 1492  
Qy 1367 CCGCGTGGCGGCGATGACCGCGCTGCGGAGCGGCGCTGCTGAGTTCCCGAACTGT 1426  
Db 1493 GCGCTGGGAGACCGTCAACCGCGCGCTTACGCGCGCTTACCTGAGTTCGCGGAGGCA 1552  
Qy 1427 CGGTGTTGAGCGCTGCGCGGAGCGGTTCTACGATCGGCGCGGAGCGGCGGCTGAGT 1486  
Db 1553 CCGTGTGAGCTTCTTACCGGAAGGTTCTTGCGATTCGCGGAGCGGCTTGGCGGAGT 1612  
Qy 1487 CCAAGCCATCAAGGAGACCGCGCGCGGAGGCGACCGCGAGAGAGAGAGCGGCTCC 1546  
Db 1613 CCAAACTCATAGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTA 1672  
Qy 1547 GCGCGGACCTGCGCGGCGGAGAGAGACCGGCGCGAGAACCTGATGATGCTGACCTG 1606  
Db 1673 AGGCTCTCCGCGCGCGCGGAGAGAGACCGGAGAGAACCTGATGATGCTGACCTG 1732  
Qy 1607 TCGGAGAGACCTTCAACAGCTGCTGCGGATGAGGCTTCCAGCTGCGCGCGCTTTCG 1666  
Db 1733 TCGCAAGGACCTTGGGAGGCTGCGGATGAGGCTTCCAGCTGACCGGAGCTGTTG 1792  
Qy 1667 AGGTGAGACTTACGCGCGCGTGCACAGCTGATGTCACATCCGCGGAGCGGCTGCGG 1726  
Db 1793 AGGTGAGACTTACGCGCACCTGTCACAGCTGTCAGACAGGTCGCGGCGCGCTGCGG 1852  
Qy 1727 CCGGACCAAGACCGCGCGCTGTAAGCGCGCTTCCCGGAGGCTTCATGACCGGCG 1786  
Db 1853 CCGACGCTTCCGCGCGCGCGCGGATGAGGCGGCTTCCCGGAGGATGATGACCGGCG 1912  
Qy 1787 CCGCGAAGAGGAGACCAAGAGATGATGACCGCTGGAAGAGGCGCGCGGCGGCT 1846  
Db 1913 CCGCGAAGGTCGACCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1972  
Qy 1847 ACTCGGAGGCGCTCGGATGTTTCGCGCTGAGGCGCGCGGAGCTCAGATGCTATCC 1906  
Db 1973 ACTCGGCGCGCTGCGGCTTCTGCGCTGAGGCGCGCGGCGGAGCTCAGATGCTATCC 2032  
Qy 1907 GCACATGCTGCTGCGCGGAGCGGAGTTCGCGGCTGCGGAGGCGGATGCTGCTCC 1966  
Db 2033 GCACATGCTGCGCGGAGCGGAGGCGGAGGCGGAGGCGGAGGCGGCGGCGGCGG 2092  
Qy 1967 TCTCGGAGGAGGAGGAGTTCAGCGAGACCGTGTAAAGCGCGGCGGAGTGTGACCG 2026  
Db 2093 TGTTCGAGCGGAGCGAGGATTCGCGGAAATGCTCTCAAGGCGGAGACCTGCGCG 2152  
Qy 2027 CCT 2030  
Db 2153 CCT 2156

RESULT 2  
US-09-987-614A-14  
; Sequence 14, Application US/09987614A  
; Patent No. 683382  
; GENERAL INFORMATION:  
; APPLICANT: BLANC, Veronique  
; APPLICANT: THIBAUT, Denis  
; APPLICANT: BAMA-JACQUES, Nathalie  
; APPLICANT: BLANCHE, Francis  
; APPLICANT: COUZET, Joel  
; APPLICANT: BARRIERE, Jean-Claude  
; APPLICANT: DEBUSCHE, Laurent  
; APPLICANT: FAMECHON, Alain  
; APPLICANT: PARIS, Jean-Marc  
; APPLICANT: DUTRUC-ROSSET, Gilles  
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By  
; FILE REFERENCE: Streptogramin genes  
; CURRENT APPLICATION NUMBER: US/09/987, 614A



CURRENT FILING DATE: 2001-11-15  
PRIOR APPLICATION NUMBER: US/08/765,907  
PRIOR FILING DATE: 1997-03-20  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 14  
LENGTH: 2220  
TYPE: DNA  
ORGANISM: Streptomyces prietinaespiralis  
US-09-987-614A-14

Query Match 39.8%; Score 820; DB 4; Length 2220;  
Best Local Similarity 64.4%; Pred. No. 1,76-126;  
Matches 1356; Conservative 0; Mismatches 655; Indels 93; Gaps 5;

QY 2 TGCGCAGCTTCTGATGAGCAACTAGCTGCTTACCCCAACCTTTCAGTACACTG 61  
DB 71 TCGCAACCTTGCTATGACACTAGCTGCTTACCTTACCTTCCAGATGCTGG 130

QY 62 GCGAGCCACCGGCGCAACCCCGCTGCTGTCGCAAGACGCGGACTGTGCGGCTGC 121  
DB 131 CCGAGGTGAAGCGGCGCGCTCCGCTGCTGCTCCGCAAGACGACACCCGCACTGGCAG 190

QY 122 CCGTCGAG-----GACTTGAAGCGATGCTGCTGCTCCCGGAGCCCGGCGAGCC 175  
DB 191 CCCTGCGCGCGGCGACTTGCACAAAGTGTGCTGTCAACCCGCGCCGCGCACCCGCA 250

QY 176 GCGAAGCGGACTTGGGAATCAGCCGCGGCGGATCAGCAGAGGCGCTGCGCTCG 235  
DB 251 CCGACACCGACTGGGCTTACGCGCGCGGTGATCAGCAATGGGAACCTGCGCTGCTG 310

QY 236 GCGTCTGCTGCGGCGACGAGGCGATCGCGGCTTCTGCGGGAACCGTGGCGCTGCGCC 295  
DB 311 GGGTGTGCTGCGGCGACGAGGCGCTGTGCTGCTGCGCGCGCGCGCTGCTCAACGAC 370

QY 296 CCGAAGCCATGACAGCGCGGCTCTCGAGGTGCGGACACCGGCGAGAGACTTTCGCG 355  
DB 371 CCGAAGCCCTTTCAGCGGCGGACGACGACATCGCGCAAGAGGCGGCGCTTTCGCGA 430

QY 356 GCGTCCCTGCGCGCTTACCGCGCGTGGCTTACCTCCCTGCGCGGCGACGACTCCCGC 415  
DB 431 ACATCCCTTCCCGCTACCGTGTGCTGCTTACCTGCTGACCGTCCGCAACTGCGCG 490

QY 416 ACAGCTCGAAGCCCTGCGCTGAGCGACGAGCGGAGTGTATGGGCTGCGGCGCGG 475  
DB 491 CCGAAGCTGCGGCGACCGCGCACCGCGAGCGGCGAGCTGATGGCGCTGCGCTACGCG 550

QY 476 AGAAGCGCTGTGGGCGCTCAAGTTCACCGGAGTCCATCGGCGAGGACTTTCGCGCG 535  
DB 551 ACCTGCGCGCTTGGGCTGAGTTCACCGCGAATGATCAGACGGAACCGGCGCAC 610

QY 536 AGATCATGCGCACTTTCGCGACT----- 560  
DB 611 GATGCTCGGCACTTTCGCGACTGTCTGCGCGCGCGCGCGACCGCCCCCGACA 670

QY 561 -----CGCCCTCGCGCACACCGGCGACGCGCGCTTCCGCGCGCGCGCGCGA 604  
DB 671 CCGAAGCATACCGCGACCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 730

QY 605 -----CGTACGAACTTTCAGCGGCGCGGCTGAGCTGCTGCGCGGAG 646  
DB 731 CGTCCGCGCGGCGGAGATACCGGCTGATGATGCGCGGAGTGTGCTGCTGCGCGAG 790

QY 647 CCGAAGAGTACCGCGGCGCTGCTGCGCGGAGGAGACACGTTTGTGCTGAGACGA 706  
DB 791 CCGAGCGCGGCTTACCGCGCTGCTGCGCGCGCGCGCGCGCGCGCTTGTGCTGAGACGA 850

QY 707 GCTCCGCTCTGAGAGGCGCTGCGGCTTCTTCTCTGCGGAGACGCGCGCGCGCTG 766  
DB 851 GCGCGTGAAGCGCGGCGCTGCGCGCTTCTTCTCTGCGGCGCGCGCGCGCGCTG 910

QY 767 CCGAGTACTTACCTACCGGCTGCGCGAGCGGCTGCTGCTGCGCGGCTCGGCGGCA 826  
DB 767 CCGAGTACTTACCTACCGGCTGCGCGAGCGGCTGCTGCTGCGCGGCTCGGCGGCA 826

DB 911 GCGAAGATCACTACGATGAGCGTGCAGACCGGCGCGTGCCTGCAAGAGACGTTCAAGCG 970  
QY 827 CCAAGACCGGAGCGCGCGCGCTTCTTCAACTACGAGAGGAGAGCTGAAAGCGCGAC 886  
DB 971 GCGAGACCGCGCGCGCGCGCGCTTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1027

QY 887 GGGTCCCGCTGCGCGCGCGAGCTGCGGAGTTCAGCTGAGCTGAGCTGAGCTGAGCTG 946  
DB 1028 GCGCGCGCGCGCGCGCGCGCGCTGCGGAGTTCAGAGCTGAGCTGAGCTGAGCTGAGCTG 1087

QY 947 GCTACGAGCTGAAGGCGAGAGACCAAGCGGCGACCGCGCGAGCGGCTCCCGGCGCGAG 1006  
DB 1088 GCTACGAGAGCAAGGCGGAGAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1147

QY 1007 CCGGTTCCCTTTCGCGCGCGCGAGCGGCGAGTGCCTTCAAGAGAGAGAGAGAGAGAGAG 1066  
DB 1148 GCGCTTCACTGCTGCGCGAG 1207

QY 1067 TGTGAGCTTTCAGACCGCGCGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1126  
DB 1208 TCTTGGAGCTGAG 1252

QY 1127 CCGAAGCTTTCAGCGGCTGAGCGGCTGCGCGCGCGCGCGAGCGAGCGAGCGAGCGAGCT 1186  
DB 1253 CCGAAGCTTTCAGCGGCTGAGCGGCGGCGCGCGCGCGCGCGAGCGAGCGAGCGAGCGAGCT 1312

QY 1187 TCGGAGTCCCGAGGCGGCGGCGGCTTGGGCGCGCGCGCGCGCGCGCGCGCGCGAGAG 1246  
DB 1313 TCAACCTGCGCGCGAG 1372

QY 1247 ACAGCTTACCTCAAGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1306  
DB 1373 CCGGCTTACCGGAG 1432

QY 1307 AGATGCTGAGCAAG 1366  
DB 1433 AGATGCTGAGCAAG 1492

QY 1367 CCGGCTGAG 1426  
DB 1493 GCGCGCTGAG 1552

QY 1427 CCGGCTGAG 1486  
DB 1553 CCGGCTTACCTTCAAG 1612

QY 1487 CCAAGCCATCAAG 1546  
DB 1613 CCAAGCCATCAAG 1672

QY 1547 GCGCGAGCTGAGCGGCGGAG 1606  
DB 1673 AGGCTTCCCTGCGCGCGCGAG 1732

QY 1607 TCGGCAAG 1666  
DB 1733 TCGGCAAG 1792

QY 1667 AGGTGAG 1726  
DB 1793 AGGTGAG 1852

QY 1727 CCGGCAAG 1786  
DB 1853 CCGGCAAG 1912

QY 1787 CCGGCAAG 1846  
DB 1913 CCGGCAAG 1972

QY 1847 ACTCGGAG 1906  
DB 1973 ACTCGGAG 2032



```
QY 1907 GCAACATGCTGCTGCGCCGAGCCGAGATTGCGCGTCCGCGGGCGATGTCCTCC 1966
      |||||
Db 2033 GCACATCTGTGCGCACCGAGAGCGCGCACCATCGCGGTGGGGCGCGCTGTCGCC 2092
QY 1967 TCTCCCAACCGAGAGAGAGATTCAACCGAGACCGTGTGAAGGCCCGGCATATGTCA 2026
      |||||
Db 2093 TGTCCGACCCCGAGCAGAGAGTCCGGAATGCTCTCTCAAGGCGCAGACCACTCCG 2152
QY 2027 CCCT 2030
      |||||
Db 2153 CCCT 2156

RESULT 3
US-08-765-907A-6
; Sequence 6, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAWAS-JACOUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSCH, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSER, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Mutasynthesis
; CURRENT APPLICATION NUMBER: US/08/765,907A
; NUMBER OF SEQ. ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 4496
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-6

Query Match 22.3%; Score 459; DB 3; Length 4496;
Best Local Similarity 60.5%; Pred. No. 4.2e-67;
Matches 899; Conservative 0; Mismatches 495; Indels 93; Gaps 5;

QY 2 TGGCGACGCTTGTGATGGAACATACGACTGCTTCAACCCCAACCTGTTCAATACATCG 61
Db 3028 TGGGAACCTGTGATGGAACATACGACTGCTTCAACCTTTCACATGCTGG 3087
QY 62 GCGAGGCAACCGGAGCAACCCCGTGTGTCGCCAAGACGCGACTGTGCGGGCTG 121
      |||||
Db 3088 CCGAGGGAACCGGCGCGCTGCTGCTGCTCGGCAAGACGACACCCGACCTGGCAG 3147
QY 122 CCGTCAAG-----GACTTGAAGCGATGCTGTGTCGCCGAGCCCGGCGAGCCGAC 175
      |||||
Db 3148 CCGTGGCGCGGAGACTTGAACAACTGCTGTCAACCGGCGCGGCGACCCGCCA 3207
QY 176 GCGAAGCGGACTTGGAAATCAAGCCGCGGCGATCAACGACAGCGGCTTCCGCTCG 235
      |||||
Db 3208 CCGACACGACGTGGGCTCAGCGCGGAGTGAATCAACGAAATGGAACTTCCCTGTC 3267
QY 236 GCGTCTGCTCGGCAACGAGGCAATCGCCAGCTCTTCCGCGGAACGTCGAGCTCGCC 295
      |||||
Db 3268 GGGTGTCTGTGGGCAACGAGCCCTGTGCTGCTGCGCGCGCGCGCTGTCAAGCAC 3327
QY 296 CCGAACCATGCAACGCGCGGAGTCTCGAGGTGCGGCAACCGCGGAGAGAGTCTTCGG 355
      |||||
Db 3328 CCGAACCTTTCAAGCGCGGACCAAGCAATCCGCCAAGAGGAGGCTGTTCGCGA 3387
QY 356 GCGTCTGCTGCGGCTTACCGCGGTGCGTACCACTCTGCGCGGCGACCACTCCCG 415
      |||||
Db 3388 ACAATCCCTCCCGCTGACCGTGTGCTACCACTGCTGACGCTCCGCAACTGCGCG 3447
```

```
QY 416 ACAGCTCGAACCCCTCGCTTGAAGGAGCAACGAGGTCTGTATGAGGCTTCGAGCACCG 475
      |||||
Db 3448 CCGACTGCGCGCAACCGCCCAACCGCGGAGCGGAGCTGATGCGCGTCCGACCGCC 3507
QY 476 AGAAGCGCTGTGGGCGTTCAGATTCAACCGGAGTTCATTCGAGACGCACTTCGGCGGG 535
      |||||
Db 3508 ACCTGCGCGCTTCCGCGCGGAGTTCACCCGAAATCGATCAGACGCGAACCGGCGAC 3567
QY 536 AGATCATGGCAACTTCCGAGACT----- 560
      |||||
Db 3568 GGAATGCTCGCAACTTCCGAGACTGTCTCTGCGCGGCGCGGACCGCCCGCCGACA 3627
      |||||
QY 561 -----CGCCCTGCGCCACCGCGGAGCAACGAGCGCGCAACGAGGCGCATCTCCC 605
      |||||
Db 3628 CCGAAGCATACCGGACCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3687
      |||||
QY 606 -----GTACGAATCTCAAGTGGCGCGCTCGACGTCGTCGGAGAC 646
      |||||
Db 3688 CGTCCGCGCGGAGGAGTACCGGCTGCAATGTGCGCAGAGTGTGCTGCTGCGCGAG 3747
      |||||
QY 647 CCGAAGAGGTAGCGCGCGCTGCTGCGCGGAGGAGCAACGATTGCTGAGCAGCA 706
      |||||
Db 3748 CGAATCCGCGTTACCGCGCTGTGCGCGAGCGCGCGCGCGCTGCTGAGTTCAGACGA 3807
      |||||
QY 707 GCTTCGCTCGAAGCGCGCTGCGGCTTCTCTCGCGCAGCAACCGCGCGCGCTTG 766
      |||||
Db 3808 GCGCGGTGAGCGGCGCTGCGCGCTTCACTTCTGCGCGCGCGCGCGCGCGCGCTG 3867
      |||||
QY 767 CCGAGTACTCACTACCGCGCTGCGCGAGCGGCTGTCTCTGCTCGCGCGCTCGACGGA 826
      |||||
Db 3868 GCGAAGATCACTACCTACGATGCGCGAGCGGCGGTGCGCTCAAGAGCGGTTAAGG 3927
      |||||
QY 827 CCAAGACCGGAGCGCGCGCGCTTCAATACCTGAGAGGAGCGCTGAAACGCGAG 886
      |||||
Db 3928 GCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3984
      |||||
QY 887 GGGTCCCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 946
      |||||
Db 3985 GCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4044
      |||||
QY 947 GCTAAGCTGAAGGCGGAGCAACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1006
      |||||
Db 4045 GCTAAGCTGAAGGCGGAGCAACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 4104
      |||||
QY 1007 CGCGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1066
      |||||
Db 4105 GCGCTTCAATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4164
      |||||
QY 1067 TGTGCGCTTGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1126
      |||||
Db 4165 TCTGTGCACTGAGGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4224
      |||||
QY 1127 CCGAGACCTTCAACCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1186
      |||||
Db 4225 CCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4284
      |||||
QY 1187 TGGGATTCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1246
      |||||
Db 4285 ACACCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4329
      |||||
QY 1247 ACAGCTTCAAGGCGATGCAAGTGTCTCAAGAGATTCGCAACCGCGAGTGTAG 1306
      |||||
Db 4330 CCGCTACCGGAGACGTGTGAGAAATGCGCGCGCTGTATCAACGAGCGGAGACTAG 4389
      |||||
QY 1307 AGATGCTTGAACCAATGTATCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCTTACT 1366
      |||||
Db 4390 AGGTGTCTGAAGCAATGTCTCGGAGTGCAGCGCGCGAGTCAACCGCGCTAC 4449
      |||||
QY 1367 CGCGGTGCGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1413
      |||||
Db 4450 GCGCGTGGCAACGTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4496
```



RESULT 4  
US-09-987-614A-6  
; Sequence 6, Application US/09987614A  
; Patent No. 683382  
; GENERAL INFORMATION:  
; APPLICANT: BLANC, Veronique  
; APPLICANT: THIBAUT, Denis  
; APPLICANT: BAMES-JACQUES, Nathalie  
; APPLICANT: BLANCHE, Francis  
; APPLICANT: COUZET, Joel  
; APPLICANT: BARRIERE, Jean-Claude  
; APPLICANT: DEBUSCHE, Laurent  
; APPLICANT: FAMECHON, Alain  
; APPLICANT: PARIS, Jean-Marc  
; APPLICANT: DUTRUC-ROSET, Gilles  
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By  
; FILE REFERENCE: Streptogramin genes  
; CURRENT APPLICATION NUMBER: US/09/987,614A  
; PRIOR FILING DATE: 1997-03-20  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 4496  
; TYPE: DNA  
; ORGANISM: Streptomyces pristinaespiralis  
US-09-987-614A-6

Query Match 22.3%; Score 459; DB 4; Length 4496;  
Best Local Similarity 60.5%; Pred. No. 4.2e-67;

Matches 899; Conservative 0; Mismatches 495; Indels 93; Gaps 5;

QY 2 TGGCAGCGCTTGTGATGCACTAGCACTGTTCAACCACTGTTCCAGTACATCG 61  
DB 3028 TGGGAACCGTGTGATGCACTAGCACTGTTCAACCACTGTTCCAGTACATCG 3087  
QY 62 GCGAGCCACCGGGCAACCCCGTGTGTGCTGCCAAGCGCGGACTGTGCGGCTCG 121  
DB 3088 CCGAGGTGAACGCGCGCTCCCTGTGTGTGCGCAACGACCCGCACTGGCAGG 3147  
QY 122 CGGTGAG-----GACTTGAAGCGATGTGTGTCCTCCCGGGCCCGGAGCCGAC 175  
DB 3148 CCTTGGCGCGGGCGACTTGAACGTCGTGTCTTCAACCGGCGCGGCGACCCGCA 3207  
QY 176 GGGAGCGGACTTGGAAATCAGCGCGGCGATCAACGAGCGGCTGCGCTCG 235  
DB 3208 CCGACACCGACTGTGGGCTTACGCGCGGCTGATCAACGAAATGGAGCTGCGCTCG 3267  
QY 236 GCGTGTGCTGTGCGCAACGAGGATGCGCCAGCTTTGCGCGGAACTGTGCGCTCG 295  
DB 3268 GGGTGTGCTGTGGGCAACGAGGCTGTGTGCTGTGCGCGGCGCGCTGTGTCAGCAG 3327  
QY 296 CCGAACCCTATGCAACGCGCGGCTTCCGAGGTGCGGCAACCGCGGAGAGTCTTCCG 355  
DB 3328 CCGAACCCTTTTCAACGCGCGGCAACGAGCAATCGGCAACGCGGAGGCTGTTCGCA 3387  
QY 356 GCGTGTGCTGTGCGCTTCAACGCGCGGCTTACCACTTCCGCGGCAACCGGCTCG 415  
DB 3388 ACATCCCTTCCGCTGACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3447  
QY 416 ACGAGCTGAACCCCTGTGCTGAGGCGAAGCGGAGTGTGATGAGCTTGGCGACCGG 475  
DB 3448 CCGAGCTGTGCGCGCAACCGCAACGCGGAGGCTGATGAGCGCTGCGCCACCGCG 3507  
QY 476 AGAAGCGCTGTGCGGCGCTTCAATCCCGGAGTCACTGCGGAGGACTTGGCGG 535  
DB 3508 ACGTGTGCTGTGCGGCTGCGGAGTTCACCCCGAATGATCAACGAGCAACGCGGCG 3567  
QY 536 AGATCATGGCAACTTCCGCGACT----- 560

DB 3568 GATGCTGTGCAACTTCCGCGACTGTCCCTGTGCGCGGCGCGGCAACCGGCGCGGCA 3627  
QY 561 -----CGCTTGTGCGCGGCAACCGGCGGCAACCGGCGGCGCGGCGGCGGCGG 605  
DB 3628 CCGAAGCATACCGCGACCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGGCGGCGG 3687  
QY 606 -----GTACGAACCTCAACGTTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 646  
DB 3688 CGTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3747  
QY 647 CCGAAGAGTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 706  
DB 3748 CCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3807  
QY 707 GCTTGTGCTTCCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 766  
DB 3808 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3867  
QY 767 CCGAGTACCTCACTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 826  
DB 3868 GCGAAGCATACCTTCAACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3927  
QY 827 CCAAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 886  
DB 3928 GCGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3984  
QY 887 GGGTCCCGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 946  
DB 3985 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4044  
QY 947 GCTACGAGCTGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1006  
DB 4045 GCTACGAGCTGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4104  
QY 1007 CCGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1066  
DB 4105 GGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 4164  
QY 1067 TGTGTGCTTCAACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1126  
DB 4165 TCTTGTGCACTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 4224  
QY 1127 CCGAAGCCTTCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1186  
DB 4225 CCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4284  
QY 1187 TCGGAGTCCCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1246  
DB 4285 ACGACCACTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4329  
QY 1247 ACGCTTACTTCAAGCGCATGCAAGTGTCTTCAAGAGATTCGCAACGCGGAGTGTGAG 1306  
DB 4330 CCGCTTCAACGCGGAGCTGTGAGGATGCGCGGCTGATTCACGAGCGGCGGAGCTAG 4389  
QY 1307 AGATGTGCTTCAACGAGTGTGCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1366  
DB 4390 AGGTGTGCTTCAACGAGTGTGCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4449  
QY 1367 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1413  
DB 4450 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4496

RESULT 5  
US-08-765-907A-1  
; Sequence 1, Application US/08765907A  
; Patent No. 6352839  
; GENERAL INFORMATION:  
; APPLICANT: BLANC, Veronique  
; APPLICANT: THIBAUT, Denis  
; APPLICANT: BAMES-JACQUES, Nathalie  
; APPLICANT: BLANCHE, Francis  
; APPLICANT: COUZET, Joel



```

; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSETT, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; TITLE OF INVENTION: Mutasynthesis
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/08/765, 907A
; CURRENT FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;
; LENGTH: 2888
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-1

Query Match: 17.7%; Score 365.4; DB 3; Length 2888;
Best Local Similarity 74.2%; Fred. No. 1e-51;
Matches 462; Conservative 0; Mismatches 161; Indels 0; Gaps 0;
```

	Query Match	17.7%;	Score 365.4;	DB 3;	Length 2888;
	Best Local Similarity	74.2%;	Pred. No. 1e-51;		
	Matches 462;	Conservative 0;	Mismatches 161;	Indels 0;	Gaps 0;
0y	1408 CTCGAGTCCCCGGAACGTGCTGGTGTGAAGCCCTCGCCCGAGCGTTTCTCACAATCGGC 1466				

QY	1408	CTCGAGTTCCTCCCGAACTGTGGTGTGAGGCGCTGACCCGAGCGGTTCTCAGATCGGC	1467
Db	1	CTGCAGTTCCTCCCGAGGCCACCGCTGTCACTCTCACCAGACGGTTCTTGCGATCGGC	60
QY	1468	GCCGACGGCGGCGTTCGAGTCCGAAGCCATCAAGGGAGACCCGCCCCCGGGGGCGGACCGCG	1527
Db	61	GGCGACGGCTGGGGGAGTCCAAACCCATCAAGGGCACCCGCCCCCGGGGGCGGGCCCC	120
QY	1528	GAGAGGACCGAGCGGCTTCGCGCCGCACTGGCCGGCCGGGAGAAAGACCCGGAGC	1587
Db	121	GCCGAGGACGGCGCGCTTCAGGCTTCCTTCGCGCGGGCGGAGAAAGACCGGAGGAGAAC	180
QY	1588	CTGATGATCGTCGACCTGTGTCGCAACGACCTCAACAGCTGTGGCGAGTCGGCTCGCTC	1647
Db	181	CTGATGATCGTCGACCTGTGTCGCAACGACCTTCGGCCAGGCTCTCGACATGCGCTCGCTC	240
QY	1648	CACGTGCCCGGCGCTCTTCGAGGTGAGAGCTTACGCGCCCTGTGCACCAAGCTGTGTGAC	1707
Db	241	CACGTACCGGGCGCTGTTGTGAGGTGAGAGCTTACGCCACCTGTGCACAGCTGTGACGACG	300
QY	1708	ATCCGGGGGAGGGCTGCGGCGCGGGACCGACGACCGCGCGTGTGATGCGCGGCTTCCCC	1767
Db	301	GTCCGCGGCGCGCTGTGGCGGGCGGAGTCTCCGCGCCCGCGGTACGGGCGGCTTCCCC	360
QY	1768	GGCGGCTTCATGACCGCGCGCGCCCAAGAGCGCACATGAGATCATGACCGGCTGGAG	1827
Db	361	GGCGGGTGTGATGACCGGGCGGCGCCCAAGGTCCGCACGACGATTCATGACCGGCTCGAG	420
QY	1828	GAAAGCCCCCGGGGGCTTACTCCGGGGGCGCTCGATGTTGTCGCCCTCAGCGGCGGCC	1887
Db	421	AAGGGCCCCCGCGCGGTGTACTCTCGGGCGCGCTGTGGCTTACTTCGCGCTCAGCGGCGGCC	480
QY	1888	GACCTCAGCATCGTCATCCGACCATCGTGTGGCCGACGGGCGGAGATTGTGGGCTC	1947
Db	481	GACCTCAGCATCGTCATCCGACCATCGTGTGGCCACCGAGGGGCGGCCACCATGGGCTG	540
QY	1948	GGCGGGGCGATCGTGTCCCTCTTCGACCCGAGAGGAGGATTCACCGAGACCGTGTAAAG	2007
Db	541	GGCGGGCGCGTGTGCGCCTGTCCGACCCCGACGACGAGTGTCCGGAATGCTCCTCAAG	600
QY	2008	GCCCCGCGCATGTGTACCGGCTT	2030
Db	601	GCGCAGACCACTTCGCGGCTT	623

RESULT 6  
US-09-987-614A-1  
; Sequence 1, Application US/09987614A  
; Patent No. 683382  
; GENERAL INFORMATION:  
; APPLICANT: BLANC, Veronique

```

: APPLICANT: THIBAUT, Denis
: APPLICANT: BAMA-S-JACQUES, Nathalie
: APPLICANT: BLANCHE, Francis
: APPLICANT: COUZET, Joel
: APPLICANT: BARRIERE, Jean-Claude
: APPLICANT: DEBUSCHE, Laurent
: APPLICANT: FAMECHON, Alain
: APPLICANT: PARIS, Jean-Marc
: APPLICANT: DUTRUC-ROSSET, Gilles
: TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
: TITLE OF INVENTION: Mutasynthesis
: FILE REFERENCE: Streptogramin genes
: CURRENT APPLICATION NUMBER: US/09/987,614A
: CURRENT FILING DATE: 2001-11-15
: PRIOR APPLICATION NUMBER: US/08/765,907
: PRIOR FILING DATE: 1997-03-20
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO: 1
: LENGTH: 2888
: TYPE: DNA
: ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-1

Query Match          17.7%   Score 365.4; DB 4; Length 2888;
Best Local Similarity 74.2%; Pred. No. 1e-51;
Matches 462; Conservative 0; Mismatches 161; Indels 0; Gaps 0

```

ORGANISM: Streptomyces pristinaespiralis  
US-09-987-614A-1

	Query Match	17.7%	Score 365.4	DB 4	Length 2888	
	Best Local Similarity	74.2%	Pred. No. 1e-51			
	Matches 462	Conservative 0	Mismatches 161	Indels 0	Gaps 0	
Qy	1408	CTCAGATTCCCGCAACTGTGCTGCTGAGCGCCTTGCCGACCGGTTCTTCACATCGGC	1467			
Db	1	CTGCAGATTCCCGGGGGCCACCGTGTCTGCTCTCTCACCGAAAGGTTCTCTGGCATCGGC	60			
Qy	1468	GCGACGGCGGGCGTTCGAGTCCAAAGCCCATCAAGGGAGACCCGCCCGGGGGCGGCACCGC	1527			
Db	61	GCGAGCGGCTGGGGGAGTCCAAACCCATCAAGGGACACCGCCCCCGGGCGCCGGCCCC	120			
Qy	1528	GAGAGAGCGACGCGCTCCGCGCCGCACTTGCGCGGCCGGGAGAAAGACCGGGCCAGAAAC	1587			
Db	121	GCCACGAGACGCGCGCTCMAAGGCTCTCCCTCGCGCGCGCGGAGAAAGACCGCACGCAAGAC	180			
Qy	1588	CTGATGATCTGTTCACCTGTGTCCGCAAGCACTCTCAACAGCGTCTGGCATCGGCTTCGGTC	1644			
Db	181	CTGATGATCTGTTCACCTGTGTCCGCAACCACTTCGGCGAGTCTGGCATCGGCTTCGGTC	240			
Qy	1648	CACATGCCCCCGGCTCTTCGAGGTGAGAGACTACGCGCCGCTGACACCACTGGTGTGCAC	1707			
Db	241	CACGTACCGGGCGCTTTCGAGGTGAGAGACTACGCGCCGCTGACACCACTGTGTGCACAG	300			
Qy	1708	ATCCGGGAGCGGCTGCGGCCCGGCAACAGACCCGCGCTTCGCTACGCGCGCTTCCCC	1767			
Db	301	GTCGCGGCGCGCTTCGCGGGCGCGAGTCTCCGCGCCCGCGCGTACGCGGCGCGCTTCCCC	360			
Qy	1768	GGCGGCTTCATGACCGGGCGGCGCCCAAGAAAGGACCAATGAGATCATGACCGCGCTGGAG	182			
Db	361	GGCGGCTTCATGACCGGGCGGCGCCCAAGGTCTCGCACATCATGATTCATGACCGGCTCGAG	420			
Qy	1828	GAAAGCCCCCGGAGGCGTCTACTCCGAGGCGGCTCGAGATGTTTCGCGCTTCAGCGCGCGCC	1887			
Db	421	AAGGCGCCGCGGCGGTACTCTCGGAGCGCGCTTGAGGCTTACTTCGCGCTTCAGCGCGCGCC	480			
Qy	1888	GACCTTCAGCATTCGTTCATTCGCAACCATCGTGTGGCCCAACGGCCAGGCGGAATTCGGCGTC	1944			
Db	481	GACCTTCAGCATTCGTTCATTCGCAACCATCGTGTGGCCCAACGGCGGCGCCACATTCGGCGTC	540			
Qy	1948	GCGGAGCGATCTGTCCCTCTTCGACACGAGAGAGAGTTTCACGAGACCGTGGTAAAG	2007			
Db	541	GCGGAGCGCGTCTGTCCCTCTTCGACACCGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGG	600			
Qy	2008	GCCCGCGCATGTTCACCGCGCTT 2030				
Db	601	GCGCAGACCACTTCGCGCGCGCTT 623				



RESULT 7  
US-08-403-852D-9  
Sequence 9, Application US/08403852D  
Patent No. 5891695  
GENERAL INFORMATION:  
APPLICANT: Blanc, Veronique  
APPLICANT: Blanche, Francis  
APPLICANT: Crouzet, Joel  
APPLICANT: Jacques, Nathalie  
APPLICANT: Lacroix, Patricia  
APPLICANT: Thibaut, Denis  
APPLICANT: Zagorec, Monique  
APPLICANT: Debussche, Laurent  
APPLICANT: De Crey-Lagard, Valerie  
TITLE OF INVENTION: Polypeptides Involved In The  
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fimegan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/403,852D  
FILING DATE: 10-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR 93/00923  
FILING DATE: 25-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/11441  
FILING DATE: 25-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyer, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03806.0054-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4000  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 645 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: S.pristinaespiralis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 61..645  
OTHER INFORMATION: /product= "gene papa"  
US-08-403-852D-9  
Query Match 13.2%; Score 272.2; DB 2; Length 645;  
Best Local Similarity 68.7%; Pred. No. 2.2e-36;  
Matches 392; Conservative 0; Mismatches 173; Indels 6; Gaps 1;  
QY 2 TGGGACGCTTTCGATGAGAACTAGACACTCGTTACCCCAACCTGTTCCAGTACATCG 61  
DB 71 TGGGACGCTTTCGATGAGAACTAGACACTCGTTACCCCAACCTGTTCCAGTACATCG 130  
QY 62 GCGAGCGCAACCGGCAACCCCGCTGCTGCGCAACGAGCGGCACTGGTCCGGCTGC 121

DB 131 CCGAGTGTAAGGCGCGCTCCGCTCGTCGCGCAACGACGACACCCGCACTGGCAGG 190  
QY 122 CCGTCGAG-----GATTTGACGCGATCTCTGTCGTCCCGGGCCCGGCAAGCCCGAC 175  
DB 191 CCGTGGCGCGGGGAGATTGACAACTCTCTCTCAACCCGGCCCGGCAACCCCGCA 250  
QY 176 GGGAAACGGACTTCGGAATCAGCGCGCGGCGATGACGACGAGCGGCTGCTGCTCG 235  
DB 251 CCGACACCGACCTTGAGGCTTACGCGCGGAGTATGACCGAATGAGACCTTCCGCTCG 310  
QY 236 GCGTTCGCTCGGCGCAACGAGGATCGCCAGCTCTTTCGCGGAAACCGTCGAGCTCGCC 295  
DB 311 GGGTGTCTGAGGACACGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 370  
QY 296 CCGAACCATGACAGCGCGGCTCTCGAGGTGGGACACCGGCGAGAGCTTTCGCGG 355  
DB 371 CCGAACCCTTTCAACGCGCGGACGAGGACATCGCGCACGAGCGGAGGCTGTTCGGA 430  
QY 356 GCGTCCCGCTCGCGGTTACCGCGGCTGAGCACTCCGCGCGGCGGACCGACTCCCGG 415  
DB 431 ACATCCCTCTCCCGCTGACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 490  
QY 416 ACGAGCTGAACCCCTCGCTGAGGAGGACGAGCGGAGTGTGATGAGGCTGCGGACCGCG 475  
DB 491 CCGACCTGGCGGCGCACCGCCACACGCGCGGACGAGTGTGCTGCTGCTGCTGCTGCTG 550  
QY 476 AGAAGCGCTGTGGGCGTCACTTCCACCGGAGTGCATGCGAGGACTTTCGCGCGG 535  
DB 551 ACCGCGCGGCTTTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 610  
QY 536 AGATCATGGCCACTTCCGCGAGCTGCGCT 566  
DB 611 GGATGCTGCGCACTTCCGCGAGCTGTCTT 641  
RESULT 8  
US-08-510-646B-9  
Sequence 9, Application US/08510646B  
Patent No. 6077699  
GENERAL INFORMATION:  
APPLICANT: Blanc, Veronique  
APPLICANT: Blanche, Francis  
APPLICANT: Crouzet, Joel  
APPLICANT: Jacques, Nathalie  
APPLICANT: Lacroix, Patricia  
APPLICANT: Thibaut, Denis  
APPLICANT: Zagorec, Monique  
APPLICANT: Debussche, Laurent  
APPLICANT: De Crey-Lagard, Valerie  
TITLE OF INVENTION: Polypeptides Involved In The  
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fimegan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/510,646B  
FILING DATE: 03-AUG-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/403,852  
FILING DATE: 10-MAY-1995  
PRIOR APPLICATION DATA:



```
/ APPLICATION NUMBER: PCT/FR 93/00923
/ FILING DATE: 25-SEP-1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: FR 92/11441
/ FILING DATE: 25-SEP-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Meyers, Kenneth J.
/ REGISTRATION NUMBER: 25,146
/ REFERENCE/DOCKET NUMBER: 03806.0054-01000
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (202) 408-4000
/ TELEFAX: (202) 408-4400
/ INFORMATION FOR SEQ ID NO: 9:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 645 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ ORIGINAL SOURCE:
/ ORGANISM: S.pristinaespiralis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 61..645
/ OTHER INFORMATION: /product= "gene papa"
US-08-510-646B-9

Query Match      13.2%; Score 272.2; DB 3; Length 645;
Best Local Similarity 68.7%; Pred. No. 2.2e-36;
Matches 392; Conservative 0; Mismatches 173; Indels 6; Gaps 1;
```

2 TGGGACGGCTTGATGACAACTAGACACTGCTTACCCCAACCTGTTCCAGTACATCG 61  
71 TGGGAACCTGCTGATGACAACTAGACACTGCTTACCAACCTTTCAGATGCTGG 130  
62 GCGAGGCAACCGGGGAAACCCCGTGTGTCGCAACGACGCGACTGTGCGGGTGC 121  
131 CCGAGGTGAACGGCGCGCTCGCTGTGTGCGCAACGACGACCCGCACTGGCAGG 190  
122 CCGTCCGAG-----GACTTGCAGCCGATGCTGTGTGCGGGGCGCGGAGCCCGGACC 175  
191 CCTGGGCGCGGCGGACTTGCACAGTGTGCTGTCAACCCGCGCGGCGCACCCGCCCA 250  
176 GGAACGGGACTTTCGAATAGCCGCGCGGCGGATCAACGACGCGCTGCGCTCTCG 235  
251 CCGAACCGGACCTGGGCTTACGCGCGGGGTGATCAACGAAATGGGACTTGCCTGCTCG 310  
236 GCGTCTGCTCGGCGCAACGAGGATCGCCAGCTCTTCCGCGGAAACGTCGAGCTCGCC 295  
311 GGGTGTGCTCGGCGCAACGAGGCTGTGCTGTGCTCGCGCGCGCGCGCTGTCTCAGGAC 370  
296 CCGAACCCATGACACGCGCGGGGTCTCGAGGTGCGGCAACCGGCGAGAGCGTCTTCGGG 355  
371 CCGAACCTTTCACAGGCGCGACACGAGGACATCCGCCACGAGGCGAGGGCTGTTCGCGA 430  
356 GCGTCTGCTCGGCGCTTACCGCGCGGTGCGTACCACTTCCCTGCGCGCACGACTCCCGG 415  
431 ACATCCCTCCCGCTGACCGTGTGCTGCTACCACTGCTGACCGCTCCGCGCACTGCGCGG 490  
416 ACAGAGCTGAACCCCTGCGCTGAGAGGACGAGCGGGGTGCTATGAGGCTCGCGGACCGGG 475  
491 CCGAAGCTGGGCGGCG 550  
476 AGAAGCGGCTGTGGGGGCGTCAAGTTCACCCGGAAGTTCATCGGACGAGCACTTCGCGCGG 535  
551 ACCTGCGCGCGCTTCGGGGGTGCAAGTTCACCCCGAATGATCAGAGGAGAACAGGCGCAC 610  
536 AGATCATGGCAACTTTCGCGGACTTCGCGCT 566  
611 GGATGCTCGCAACTTCGCGGACTTCGCGCT 641

```
RESULT 9
US-09-231-818-9
/ Sequence 9, Application US/09231818
/ Patent No. 6171846
/ GENERAL INFORMATION:
/ APPLICANT: Blanc, Veronique
/ APPLICANT: Blanc, Francis
/ APPLICANT: Crouzet, Joel
/ APPLICANT: Jacques, Nathalie
/ APPLICANT: Lacroix, Patricia
/ APPLICANT: Thibaut, Denis
/ APPLICANT: Zagorec, Monique
/ APPLICANT: Debussche, Laurent
/ APPLICANT: De Crecy-Lagard, Valerie
/ TITLE OF INVENTION: Polypeptides Involved in The
/ TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
/ NUMBER OF SEQUENCES: 43
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: Pinegan, Henderson, Farabow, Garrett & Dunner
/ STREET: 1300 I Street, N.W., Suite 700
/ CITY: Washington
/ STATE: D.C.
/ COUNTRY: USA
/ ZIP: 20005-3315
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/231,818
/ FILING DATE:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/08/403,852
/ FILING DATE: 10-MAY-1995
/ APPLICATION NUMBER: PCT/FR 93/00923
/ FILING DATE: 25-SEP-1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: FR 92/11441
/ FILING DATE: 25-SEP-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Meyers, Kenneth J.
/ REGISTRATION NUMBER: 25,146
/ REFERENCE/DOCKET NUMBER: 03806.0054-00000
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (202) 408-4000
/ TELEFAX: (202) 408-4400
/ INFORMATION FOR SEQ ID NO: 9:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 645 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ ORIGINAL SOURCE:
/ ORGANISM: S.pristinaespiralis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 61..645
/ OTHER INFORMATION: /product= "gene papa"
US-09-231-818-9

Query Match      13.2%; Score 272.2; DB 3; Length 645;
Best Local Similarity 68.7%; Pred. No. 2.2e-36;
Matches 392; Conservative 0; Mismatches 173; Indels 6; Gaps 1;
```

2 TGGGACGGCTTGATGACAACTAGACACTGCTTACCCCAACCTGTTCCAGTACATCG 61  
71 TGGGAACCTGCTGATGACAACTAGACACTGCTTACCAACCTTTCAGATGCTGG 130



QY 62 GCGAGGCCACCGGGCAACCCCGCTGCTGTCGCCAAGACGCGACTGTGCGGGTGC 121  
 Db 131 CCGAGGTGAACGGGCGCGCTCGGCTGCTGTCGGAAGACACCGGCACTGGGAGG 190  
 QY 122 CCGTCGAG-----GACTTCGACCGGATGCTGTGTCCCGGGCCCGGCGACCCCGACC 175  
 Db 191 CCGTGGCGCCGGGGGCACTTCGACACGCTGTGTCACCCGGCCCGGCGACCCCGCA 250  
 QY 176 GGAACGGGACTTCGGAATCAGCCGGGGGATCACCGACAGGGGCTGCGCCCTCTCG 235  
 Db 251 CCGACACCGACTTGGGCTTCAAGCGCGGGGTGATCAGGAATGGGACCTGCGCTGCTCG 310  
 QY 236 GCGTGTGCTGCGGACACGAGGACATCCGCTCTTTCGCGGAAACGTCGCGCTCGCC 295  
 Db 311 GGGTGTGCTGGGCGACACGAGCCCTGTGCTGCTGCGCGCGCGCGCTGTTCACGAC 370  
 QY 296 CCGAACCCATGACAGCGCGGGGTCTCCGAGTGGCGCACACCGGCGAGGACGTCTTCGCGG 355  
 Db 371 CCGAACCCCTTTCAGCGCGCGCACGACGACATCCGCGCACAGCGGCGAGGGCTGTTGCGGA 430  
 QY 356 GCGTCCCTCGCGCGCTTACCGCGCGTACGCTACCTCCCTGCGCGCACGACCTCCCG 415  
 Db 431 ACATCCCTCCCGCTGACGCTGTGCTGCTACCACTGCTGACCGTCCGCGCACTGCGCG 490  
 QY 416 ACAGACTGAAACCCCTGCTGAGACGACGAGGGGTCTGATGAGGCTGCGGACCGCG 475  
 Db 491 CCGACCTGGCGGCGACCGCGCACCGCGGAGGAGCTGATGAGCGCTGCGCCACCGCC 550  
 QY 476 AGAAGCGCTGTGGGGGCTCCAGTTCCACCGGAGTCCATCGGACGACACTTCGCGCGG 535  
 Db 551 ACCTGCGCGCTTGGCGGTGAGTTCCACCGCGAATGATGATGACGAGGACACGCGCAC 610  
 QY 536 AGATCATGGCCCACTTCCGCGACTCGCCCT 566  
 Db 611 GGATGCTGCGCACTTCGCGACTGTCCT 641

## RESULT 10

US-09-635-359B-9  
 Sequence 9, Application US/09635359B  
 Patent No. 6670157

## GENERAL INFORMATION:

APPLICANT: Blanc, Veronique  
 Blanche, Francis  
 Crouzet, Joel  
 Jacques, Nathalie  
 Lacroix, Patricia  
 Thibaut, Denis  
 Zagorec, Monique  
 Debussche, Laurent  
 De Crey-Lagard, Valerie

TITLE OF INVENTION: Polypeptides Involved in The  
 Biosynthesis Of Streptogramins, Nucleotide Sequences  
 Coding For These Polypeptides And Their Use

NUMBER OF SEQUENCES: 43  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
 STREET: 1300 I Street, N.W., Suite 700  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20005-3315

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/635,359B  
 FILING DATE: 09-Aug-2000  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 09/231,818

FILING DATE: 15-JAN-1999  
 APPLICATION NUMBER: US 08/403,852  
 FILING DATE: 10-MAY-1995  
 APPLICATION NUMBER: PCT/FR 93/00923  
 FILING DATE: 25-SEP-1993  
 APPLICATION NUMBER: FR 92/11441  
 FILING DATE: 25-SEP-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Meyers, Kenneth J.  
 REGISTRATION NUMBER: 25,146  
 REFERENCE/DOCKET NUMBER: 03806.0054-03000  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 408-4000  
 TELEFAX: (202) 408-4400  
 INFORMATION FOR SEQ ID NO: 9:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 645 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: S.pristinaespiralis  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 61..645  
 OTHER INFORMATION: /product= "gene papa"  
 SEQUENCE DESCRIPTION: SEQ ID NO: 9:  
 US-09-635-359B-9

Query Match 13.2%; Score 272.2; DB 4; Length 645;  
 Best Local Similarity 68.7%; Pred. No. 2.2e-36;  
 Matches 392; Conservative 0; Mismatches 173; Indels 6; Gaps 1;

QY 2 TGGGCAAGCTTGTGATGCACTAAGACTGCTTCAACCAACCTGTTCCAGTACATCG 61  
 Db 71 TGCGAACCTGTGATGCACTAAGACTGCTTCAACCAACCTGTTCCAGTACATCG 130  
 QY 62 GCGAGCCACCGGGCAACCCCGTGTGTCGCGCAACGCGCGACTGGTGGCGGTGC 121  
 Db 131 CCGAGGTGAACGGGCGCGCTCGGCTGCTGTCGCAACGACACCGGCACTTTCAGATGCTGG 190  
 QY 122 CCGTCGAG-----GACTTCGACCGGATGCTGTGTCCCGGGCCCGGCGACCCCGACC 175  
 Db 191 CCGTGGCGCCGGGGGCACTTCGACACGCTGTGTCACCCGGCCCGGCGACCCCGCA 250  
 QY 176 GGAACGGGACTTCGGAATCAGCCGGGGATCGCCAGTCCGAGCGACGAGCGCTGCGCTCTCG 235  
 Db 251 CCGACACCGACTTGGGCTTCAAGCGCGGGGTGATCAGGAATGGGACCTGCGCTCTCG 310  
 QY 236 GCGTGTGCTGCGGACACGAGGCGATCGCCAGCTTTCGCGGAAACGTCGCGCTCGCC 295  
 Db 311 GGGTGTGCTGGGCGACACGAGCCCTGTGCTGCTGCGCGCGCGCGCTGTTCACGAC 370  
 QY 296 CCGAACCCATGACAGCGCGGGGTCTCCGAGTGGCGACACCGGCGAGCGTCTTCGCGG 355  
 Db 371 CCGAACCCCTTTCAGCGCGCGACACGAGGACATCCGCGACGAGCGGCGCGCTTTCGCGA 430  
 QY 356 GCGTCCCTCGCGGCTTACCGCGCGTGTGCTTCCAGCTGCTGCGCGCGACACTTCGCGG 415  
 Db 431 ACATCCCTCCCGCTGACGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 490  
 QY 416 ACAGACTGAAACCCCTGCTGAGACGAGCGGGGTCTGATGAGGCTGCGGCGCGCGCGG 475  
 Db 491 CCGACCTGGCGGCGACCGCGCACCGCGGAGGAGCTGATGAGCGCTGCGCCACCGCC 550  
 QY 476 AGAAGCGCTGTGGGGGCTCCAGTTCCACCGGAGTCCATCGGACGACACTTCGCGCGG 535  
 Db 551 ACCTGCGCGCTTGGCGGTGAGTTCCACCGCGAATGATGATGACGAGGACACGCGCAC 610  
 QY 536 AGATCATGGCCCACTTCCGCGACTCGCCCT 566











Db	437	TTGTGCGAAGACTTTCCTCAACGACGACACCCCGACCCGGGACGCTTAAGGCCACCGCATTTG	496
Qy	1079	ACCGCCGGGGCCACGACGACGACGCGCCCGCGCTGGCTTGGCGGAGACGCGCCGAAACCTTCA	1138
Db	497	CCCGTCCGTGCGGTGACGCGACGCGACGCGCGCGCGCTGAAGAGACGCGGAGGGGCGCGCG	556
Qy	1139	CCGCGCTTGGCGCTCCGCGCCCCCGGCGGAGCCGACCCCGCCATGTGCTTTCCGGGATCCCGG	1198
Db	557	AAGCCATATGGCGGAGATG---GTTTGTGAAACCACTGGCCACCAAGAGGCGCTTATGCCCGAG	612
Qy	1199	AGGCGGCGGCGCGCTTTCGGCCCCCTTCGGCGCGCGCGCGCGCACGACAAAGAGCGCTACCTCA	1258
Db	613	CCGCGCATGTC--GTGGGAGACGGCGCGCTTACCGCGTGTGGAGACCGCGCGGGATGTAGT	670
Qy	1259	AGCGCATCGACGAGTGCCTCAAGAGATCCGCAACGCGGAGCTGTACGAGATCTGCTTGA	1318
Db	671	CGGCGGTGGCTCAGCGCGGACGAGAGTACATCCGCGCGAGACATCTTCCAGGATCGTGTCTG	730
Qy	1319	CCAACTATGGTCAACGCGCGGACCGAGCGGCGGCGCGCTTGCCTTACTTCGCGCTGCGCG	1378
Db	731	CGCGCGGCTTGCAGATCCCGGGGCGCGCGCGCGCTTGCCTTACCGCGCGCTGCGCGGCG	790
Qy	1379	CCATAGCGCCCGTCCCGTAAAGCGCGCGCTTGCAGTTCCCGGAATGTCGTCTGAGCG	1438
Db	791	GGGTGAATTCATTCGCGGTAACCTTTTCTCTGTGGAGCTGGGCGAGGCGCGCGCTGTGG	850
Qy	1439	CCTCGCCCGAGCGGTTTCTTCAACGATCGGCGCGGACCGGCGGCGGCTGAGTCCAAAGCCATCA	1498
Db	851	GCGCGTCCGCGGAGTGTGCTGTGTGACAGGTGCGCGACGAGGAGACGTGTGTGCGCGCCATCG	910
Qy	1499	AGGGAGCCCGCCCCGGGGCGGACCGCGGAGAGAGACGAGCGGCTCCGCGCGACCTTG	1558
Db	911	CCGGACACCGGCGCGGTGTGTGTGTCCGAGCGGAGACCTGAGCGCTGAGAAAGAGCTGC	970
Qy	1559	CCGCGCGGAGAAAGAACCGGGGCTGAGAACCTGATGATGTGTGACTGTGCTCCGCAACGACC	1618
Db	971	TGCGCGAAGAAAGAGAGCTCGCGGACCATCATCATGTGTGTGAGACTCGTGGGCGCAACGACG	1030
Qy	1619	TCAACAGCGTCTGCGCGATCGGCGTCCGATCGATCGTCCCGGCTCTTTCGAGGTGAGACCT	1678
Db	1031	TGGCGCGGTGCGCGCGCGCTGT	1090
Qy	1679	ACGCGCCCGGTGACCAAGCTGT	1738
Db	1091	ACAGCAGCTGATGTGACATCGT	1150
Qy	1739	CGCGCGCTGCGGTACGCGCGCGCTTTCGCCGCGCTCCATGACCGGCGCGCGCCAAAGAC	1798
Db	1151	CGCTGGATGCGGTGTGCGACCATCTTTCGCCGCGGACGAGTGTGCGGCGCGCGCCAAAGATTC	1210
Qy	1799	GCACCATGAGATTCATTCGACCGCGCTCGAGAGAAAGCCCCCGGGGGGTCTACTTCGSGGGGCG	1858
Db	1211	GCGCGATCANAATTCATTCGACGAGTTGGAGGTTCACGCGCGCGGCGCGCTTATTCGSGCGGG	1270
Qy	1859	TGCGATGAGTTGCGCTTCAGCGCGCGCGCGGACCTTACGATCGTATCCGACCAATCGTGC	1918
Db	1271	TGGGCTTACTGTCTTTCGCGGACGCTGTGACGTGTGCGTGTGCGATCTTCTTTCG	1330
Qy	1919	TGGCGGACGCGCCGCGGAGTTCGCGGTGCGCGGCGCGCGATCGTGTCTCTTCGACGAG	1978
Db	1331	TGGAACGAGAGACGCGACGATGTGTGACCTCGGCGCGGAGTGTGTGCGGACCTCGGAGCGGT	1390
Qy	1979	AGGAGGATTCACCCGAGACCGTGTGTAAAGCGCGCGCATGTGTACCGCGCTTCAGAGCA	2038
Db	1391	CGAAGAAAGCGGACGAGACGAGAACCAAGCGCGGCGCATGTGTGCGCGCTGTGCGGCTGG	1450
Qy	2039	GCGCGGTGCGCGCGCGCGCATGA 2061	
Db	1451	CGCGCGAAGAGGTGTGCGCGTGA 1473	







**This Page Blank (uspio)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 20:50:02 ; Search time 2023.38 Seconds  
(without alignments)  
7033.027 Million cell updates/sec

Title: US-10-089-514-1

Perfect score: 2061  
Sequence: 1 atcgacgacgtcttcgatcgca.....ccgtggcgggcgcccgatga 2061

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7442561 seqs, 3452328358 residues

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Published Applications NA:\*

- 1: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/prodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/prodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/prodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/prodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/prodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/prodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 14: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 15: /cgn2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/prodata/1/pubpna/US10D\_PUBCOMB.seq:\*
- 17: /cgn2\_6/prodata/1/pubpna/US10E\_PUBCOMB.seq:\*
- 18: /cgn2\_6/prodata/1/pubpna/US10F\_PUBCOMB.seq:\*
- 19: /cgn2\_6/prodata/1/pubpna/US10F\_PUBCOMB.seq:\*
- 20: /cgn2\_6/prodata/1/pubpna/US10H\_PUBCOMB.seq:\*
- 21: /cgn2\_6/prodata/1/pubpna/US10I\_PUBCOMB.seq:\*
- 22: /cgn2\_6/prodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 23: /cgn2\_6/prodata/1/pubpna/US11\_PUBCOMB.seq:\*
- 24: /cgn2\_6/prodata/1/pubpna/US11\_NEW\_PUB.seq:\*
- 25: /cgn2\_6/prodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 26: /cgn2\_6/prodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2061	100.0	2061	20	US-10-472-587-1
2	1773.8	86.1	3305	14	US-10-126-927-68
3	1772.2	86.0	12391	14	US-10-126-927-67
4	769.8	37.4	138203	21	US-10-819-386A-1
5	606.2	29.4	2196	15	US-10-156-761-1167
6	606.2	29.4	9025608	15	US-10-156-761-1
7	301.4	14.6	9025608	15	US-10-156-761-1

8	280.8	13.6	1479	15	US-10-156-761-6148	Sequence 6148, Ap
9	272.2	13.2	645	20	US-10-716-803-9	Sequence 9, Appl1
10	270	13.1	1365	17	US-10-282-122A-30135	Sequence 30135, A
11	267.2	13.0	2046	22	US-10-501-282-17	Sequence 17, Appl1
12	267.2	13.0	1754382	22	US-10-501-282-17	Sequence 6651, Ap
13	266.8	12.9	1860	9	US-09-738-626-1111	Sequence 1111, Ap
14	266.8	12.9	2005	21	US-10-494-675-25	Sequence 25, Appl1
15	266.8	12.9	3309400	9	US-09-738-626-1	Sequence 1, Appl1
16	234.8	11.4	42000	19	US-10-159-258A-1	Sequence 1, Appl1
17	234.8	11.4	63158	16	US-10-292-197A-1	Sequence 1, Appl1
18	231.6	11.2	1338	17	US-10-282-122A-32104	Sequence 32104, A
19	224.2	10.9	2523	17	US-10-282-122A-17825	Sequence 17825, A
20	220.4	10.7	1257	17	US-10-282-122A-33856	Sequence 33856, A
21	210.2	10.2	2160	14	US-10-138-927-76	Sequence 76, Appl1
22	210.2	10.2	2160	14	US-10-138-927-83	Sequence 83, Appl1
23	210.2	10.2	2160	17	US-10-430-011-76	Sequence 76, Appl1
24	210.2	10.2	2160	17	US-10-430-011-83	Sequence 83, Appl1
25	210.2	10.2	2160	17	US-10-430-011-126	Sequence 126, App
26	201.2	9.8	2199	17	US-10-430-011-122	Sequence 122, App
27	198.2	9.6	1353	17	US-10-282-122A-23903	Sequence 23903, A
28	196.4	9.5	636	15	US-10-156-761-4319	Sequence 4319, Ap
29	192.8	9.4	225646	19	US-10-470-565-1	Sequence 1, Appl1
30	190.4	9.2	1333	17	US-10-282-122A-19610	Sequence 19610, A
31	189.6	9.2	2190	14	US-10-138-927-75	Sequence 75, Appl1
32	189.6	9.2	2190	17	US-10-430-011-75	Sequence 75, Appl1
33	186	9.0	661	20	US-10-425-115-81629	Sequence 127, App
34	182.2	8.8	2166	17	US-10-430-011-127	Sequence 121, App
35	180.8	8.8	2202	17	US-10-430-011-121	Sequence 121, App
36	178.6	8.7	65140	19	US-10-203-295-1	Sequence 1, Appl1
37	178.6	8.7	125401	19	US-10-203-295-35	Sequence 35, Appl1
38	173.8	8.4	2190	14	US-10-138-927-93	Sequence 93, Appl1
39	173.8	8.4	2190	17	US-10-138-927-93	Sequence 93, Appl1
40	173.6	8.4	1119	15	US-10-156-761-6781	Sequence 6781, Ap
41	170.6	8.3	2190	14	US-10-138-927-1	Sequence 1, Appl1
42	170.6	8.3	2190	14	US-10-138-927-84	Sequence 84, Appl1
43	170.6	8.3	2190	14	US-10-138-927-85	Sequence 85, Appl1
44	170.6	8.3	2190	14	US-10-138-927-86	Sequence 86, Appl1
45	170.6	8.3	2190	14	US-10-138-927-87	Sequence 87, Appl1

#### ALIGNMENTS

RESULT 1

US-10-472-587-1

; Sequence 1, Application US/10472587

; Publication No. US20040214274A1

; GENERAL INFORMATION:

; APPLICANT: YANAI, Koji

; APPLICANT: SUMIDA, Naomi

; APPLICANT: MATANABE, Manabu

; APPLICANT: MORIYA, Tatsuki

; APPLICANT: MORAKAMI, Takeishi

; TITLE OF INVENTION: Transformants Producing Substance P/1022 Derivatives, Methods fo

; TITLE OF INVENTION: Producing The Same And Novel Biosynthesis Genes

; FILE REFERENCE: 2003-1302A/WMC/00144

; CURRENT APPLICATION NUMBER: US/10/472,587

; CURRENT FILING DATE: 2003-09-22

; PRIOR APPLICATION NUMBER: 82227/2001

; PRIOR FILING DATE: 2001-03-22

; NUMBER OF SEQ ID NOS: 40

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 2061

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(2058)

; US-10-472-587-1

Query Match 100.0%; Score 2061; DB 20; Length 2061;

Best Local Similarity 100.0%; Pred. No. 0;



Matches 2061: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGCGCAGCGCTTGTGATCGACAACTAGACTCGTTCAACCCACAACCTGTCCAGTACATC 60
Db 1 ATGCGCAGCGCTTGTGATCGACAACTAGACTCGTTCAACCCACAACCTGTCCAGTACATC 60
QY 61 GCGAGGCGCACCGGGCAACCCCGCGTGTGTGCCACGACGCGCATGTGTGCGGCTG 120
Db 61 GCGAGGCGCACCGGGCAACCCCGCGTGTGTGCCACGACGCGCATGTGTGCGGCTG 120
QY 121 CCGGTGAGGAGCTTCCAGCGGATCGTGTGTCCCGGGGCGCGGACGCGCCGACCGGGAA 180
Db 121 CCGGTGAGGAGCTTCCAGCGGATCGTGTGTCCCGGGGCGCGGACGCGCCGACCGGGAA 180
QY 181 CGGAGACTTCGGAATCAGCCCGCGGCGATCAGCAGACGCGGCTGCGCGTCTTCGCGGTC 240
Db 181 CGGAGACTTCGGAATCAGCCCGCGGCGATCAGCAGACGCGGCTGCGCGTCTTCGCGGTC 240
QY 241 TGCCTCGGCCACGAGGCGATCGCCAGCTCTTTCGGCGGAACCGTTCGGCTTCGCGCGAA 300
Db 241 TGCCTCGGCCACGAGGCGATCGCCAGCTCTTTCGGCGGAACCGTTCGGCTTCGCGCGAA 300
QY 301 CCCATGACGCGCGGGGTCTCCAGGTGCGGCGACACCGGCGAGACGTCCTCCGGGGCTTC 360
Db 301 CCCATGACGCGCGGGGTCTCCAGGTGCGGCGACACCGGCGAGACGTCCTCCGGGGCTTC 360
QY 361 CCTCGCGCTTCAACCGCGCGTGCCTACCTCCCTGGCGCGCACCGACCTCCCGACGAG 420
Db 361 CCTCGCGCTTCAACCGCGCGTGCCTACCTCCCTGGCGCGCACCGACCTCCCGACGAG 420
QY 421 CTGGAACCCCTCGCTTGAAGCGACGAGGGGTGTCTATGAGGCTTCGCGACCGCGAAG 480
Db 421 CTGGAACCCCTCGCTTGAAGCGACGAGGGGTGTCTATGAGGCTTCGCGACCGCGAAG 480
QY 481 CCGCTGTGGGGCGGTCCAGTTCCACCGCGAGTCCATCGGCGAGCATTTCCGCGCGGAGTC 540
Db 481 CCGCTGTGGGGCGGTCCAGTTCCACCGCGAGTCCATCGGCGAGCATTTCCGCGCGGAGTC 540
QY 541 ATGGCCAACTTCCGCGACTCGCCCTGCGCCACCAACCGGCGACGCGCGACCGGCGCGAC 600
Db 541 ATGGCCAACTTCCGCGACTCGCCCTGCGCCACCAACCGGCGACGCGCGACCGGCGCGAC 600
QY 601 TCCCGGTAAGAACTCCAGTGCCTGCGCGGTGACGCTGTGCGGACGCGCGAAGAGTACGC 660
Db 601 TCCCGGTAAGAACTCCAGTGCCTGCGCGGTGACGCTGTGCGGACGCGCGAAGAGTACGC 660
QY 661 CGGCGCTGCTGCCCGCGGAGGAGCACCACTTCTGTGTGACAGCACTCCGTCCTCGAA 720
Db 661 CGGCGCTGCTGCCCGCGGAGGAGCACCACTTCTGTGTGACAGCACTCCGTCCTCGAA 720
QY 721 GCGGCTCGGCTTCTCTCTCTCGGCGAGACGCGGCGCGCTCGCGCGAGTACTCACG 780
Db 721 GCGGCTCGGCTTCTCTCTCTCGGCGAGACGCGGCGCGCTCGCGCGAGTACTCACG 780
QY 781 TACCGCGTGCAGCGGCGGTGTCTCGGTCGCGGCTCCGACGCGACACAGACCGGACG 840
Db 781 TACCGCGTGCAGCGGCGGTGTCTCGGTCGCGGCTCCGACGCGACACAGACCGGACG 840
QY 841 CGGCGCGCTTCTTCAACTACCTGTGAGAGACAGTCCGACCGCGAGGATCCCCGTCGCC 900
Db 841 CGGCGCGCTTCTTCAACTACCTGTGAGAGACAGTCCGACCGCGAGGATCCCCGTCGCC 900
QY 901 CCGGAATGCGCTTTCAGGTTCAACTCGGCTACGTCGGCTTACCTTCGACTGAGCTGAG 960
Db 901 CCGGAATGCGCTTTCAGGTTCAACTCGGCTACGTCGGCTTACCTTCGACTGAGCTGAG 960
QY 961 GCGAGAGCAACCGGCGACCCCGCGCACCGGTCCCGGACCCCGACGCGCGGCTTCTTC 1020
Db 961 GCGAGAGCAACCGGCGACCCCGCGCACCGGTCCCGGACCCCGACGCGCGGCTTCTTC 1020
QY 1021 GCGGACCGGCGCATCGCTTCGACCAACAGGAGGCTGTGCTTACCTGTGCGCGCTTCGAC 1080
Db 1021 GCGGACCGGCGCATCGCTTCGACCAACAGGAGGCTGTGCTTACCTGTGCGCGCTTCGAC 1080
```

```
QY 1081 CGCGGGGCGCACGACGACCGGCGCGCGCTGTGCGGAGACGCGCCGACCTTCACC 1140
Db 1081 CGCGGGGCGCACGACGACCGGCGCGCGCTGTGCGGAGACGCGCCGACCTTCACC 1140
QY 1141 GCGCTGTGCTCGCGCGCGCGCGCGACCGCGCGACCTCCCGCATGTGTCTTGGGATTCGCGAG 1200
Db 1141 GCGCTGTGCTCGCGCGCGCGCGCGACCGCGCGACCTCCCGCATGTGTCTTGGGATTCGCGAG 1200
QY 1201 GCGGCGGCGGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260
Db 1201 GCGGCGGCGGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260
QY 1261 CGCATGACGAGTGTCTCAAGAGATCCGCAACCGCGAGTCCGACGAGATCTTGTCTGACC 1320
Db 1261 CGCATGACGAGTGTCTCAAGAGATCCGCAACCGCGAGTCCGACGAGATCTTGTCTGACC 1320
QY 1321 AACATGTGTCAACGCGCGCGACCGAGGCGACGCGCTTCGCTTACTCCGCGCTGCGCGCC 1380
Db 1321 AACATGTGTCAACGCGCGCGACCGAGGCGACGCGCTTCGCTTACTCCGCGCTGCGCGCC 1380
QY 1381 ATCAGCCCGGTCCGTAAGGCGCGCGCTGTGAGTTCCCGGAACGTGCGGTGTGAGCGCC 1440
Db 1381 ATCAGCCCGGTCCGTAAGGCGCGCGCTGTGAGTTCCCGGAACGTGCGGTGTGAGCGCC 1440
QY 1441 TCGCCGAGCGGTTCTCTCAAGATCGGCGCGCGACGCGCGCGTGTGAGTCCAAAGCCATCAAG 1500
Db 1441 TCGCCGAGCGGTTCTCTCAAGATCGGCGCGCGACGCGCGCGTGTGAGTCCAAAGCCATCAAG 1500
QY 1501 GGGACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560
Db 1501 GGGACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560
QY 1561 GCGCGGAGAGAGAACCGGCGCGAGAACCTGATGTGTGACCTGTGCGCGCGCGCGCGCG 1620
Db 1561 GCGCGGAGAGAGAACCGGCGCGAGAACCTGATGTGTGACCTGTGCGCGCGCGCGCGCG 1620
QY 1621 AACAGCGTGTGCGCGATTCGCGCTCCGTCACGTCGCGCGCGCTTTCGAGGTGAGACTTAC 1680
Db 1621 AACAGCGTGTGCGCGATTCGCGCTCCGTCACGTCGCGCGCGCTTTCGAGGTGAGACTTAC 1680
QY 1681 GCGCGCTGTGACCAAGTGTGTGTGACCACTCGGCGAGCGGCTTCGCGCGCGCGCGCG 1740
Db 1681 GCGCGCTGTGACCAAGTGTGTGTGACCACTCGGCGAGCGGCTTCGCGCGCGCGCGCG 1740
QY 1741 GCGCGCTGTGATCGGCGCGCGCTTCCCGCGCGCTCCATGACGCGGCGCGCGCGCGCG 1800
Db 1741 GCGCGCTGTGATCGGCGCGCGCTTCCCGCGCGCTCCATGACGCGGCGCGCGCGCGCG 1800
QY 1801 ACCATGAGATCATCGACCGCGCTGAGAGAGCCCGCGGCGGTCTACTCCGCGCGCGCTC 1860
Db 1801 ACCATGAGATCATCGACCGCGCTGAGAGAGCCCGCGGCGGTCTACTCCGCGCGCGCTC 1860
QY 1861 GATGTGTGTGCGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1920
Db 1861 GATGTGTGTGCGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1920
QY 1921 GCGGAGCGGCGAGGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1980
Db 1921 GCGGAGCGGCGAGGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1980
QY 1981 GAGGAGTTTACGAGAGCGTGTGTAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2040
Db 1981 GAGGAGTTTACGAGAGCGTGTGTAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2040
QY 2041 GCGGTGCGGCGCGCGCGGATGA 2061
Db 2041 GCGGTGCGGCGCGCGGATGA 2061
```

RESULT 2  
US-10-126-927-68  
; Sequence 68, Application US/10126927







```
QY 1681 GGCCTCGTGCACCACTGTTGTGTGCACCATTCGGGGAGCGCTGCGGCCCGGACACAGCAC 1740
|
|
|
Db 1630 GCGCCGCTGACAGCTGTGTGTGCACCATCCGGGGAGCGCTGCGGCCCGGACACAGCAC 1689
|
|
|
QY 1741 GCGGCTGCTTAAAGCGCGGCTTCCCGGGGCTCTCCATGACCGGGCGGCCAAGAGCGC 1800
|
|
|
Db 1690 GCGGCTGCTTAAAGCGCGGCTTCCCGGGGCTCTCCATGACCGGGCGGCCAAGAGCGC 1749
|
|
|
QY 1801 ACCATGAGATCATTCGACCCCTGAGAGAGAGCGCGGGGCGCTTACTCCGGGGCGCTC 1860
|
|
|
Db 1750 CCGATGAGATCATTCGACCCCTGAGAGAGAGCGCGGGGCGCTTACTCCGGGGCGCTC 1809
|
|
|
QY 1861 GATGTTGCTCTTCAAGCGCGGCGCGGACCTGACATCGCTCATTCGACCATCTGTCTG 1920
|
|
|
Db 1810 GATGTTGCTCTTCAAGCGCGGCGCGGACCTGACATCGCTCATTCGACCATCTGTCTG 1869
|
|
|
QY 1921 GCGGACGCGGCGGAGGCTTGGGCTGCGGGGCGGATGCTGTCTCTCTTCCGACAGAG 1980
|
|
|
Db 1870 GCGGACGCGGCGGAGGCTTGGGCTGCGGGGCGGATGCTGTCTCTCTTCCGACAGAG 1929
|
|
|
QY 1981 GAGGAGTTCAACGAGACCGTGTAAAGGCGCGGCGCATGCTCACCGGCGCTCGAGCGAGC 2040
|
|
|
Db 1930 GAGGAGTTCAAGGAGACCGTGTGTAAAGGCGCGGCGCATGCTCACCGGCGCTCGAGCGAGC 1989
|
|
|
QY 2041 GCGGTGGCGGGCGCGCGGATGA 2061
|
|
|
Db 1990 GCAGTGGCGGGCGCGCGGATGA 2010
|
|
|
```

RESULT 3  
US-10-126-927-67  
Sequence 67, Application US/10126927  
Publication No. US2003082575A1

GENERAL INFORMATION:

```
APPLICANT: The Scripps Research Institute
APPLICANT: Schultz, Peter G
APPLICANT: Wang, Lei
APPLICANT: Anderson, John C
APPLICANT: Chin, Jason
APPLICANT: Liu, David R
APPLICANT: Magliery, Thomas
APPLICANT: Meggers, Eric L
APPLICANT: Mehl, Ryan A
APPLICANT: Pasternak, Mirol
APPLICANT: Santoro, Stephen W
APPLICANT: Zhang, Zhiwen
TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
FILE REFERENCE: 54-000120US
CURRENT APPLICATION NUMBER: US/10/126,927
CURRENT FILING DATE: 2002-04-19
PRIOR APPLICATION NUMBER: US 60/285,030
PRIOR FILING DATE: 2001-04-19
PRIOR APPLICATION NUMBER: US 60/355,514
PRIOR FILING DATE: 2002-02-06
NUMBER OF SEQ ID NOS: 79
SOFTWARE: PatentIn version 3.1
SEQ ID NO 67
LENGTH: 12391
TYPE: DNA
ORANISM: Plasmid pSC101, Streptomyces venezuelae papABC
US-10-126-927-67
```

Query Match 86.0% Score 1772.2; DB 14; Length 12391;  
Best Local Similarity 92.8%; Pred. No. 0;

Matches 1912; Conservative 0; Mismatches 98; Indels 51; Gaps 3;

```
QY 1 ATGCGACGCTTGTGATGACAACTAGACTCGTTCAACCAACCTGTTCCAGTACATC 60
|
|
|
Db 24 ATGCGACGCTTGTGATGACAACTAGACTCGTTCAACCAACCTGTTCCAGTACATC 83
|
|
|
QY 61 GCGGAGGCGCACCGGGAGACCCCGGCTGTGTGTCGACGAGCGGATGTTGCGGCTG 120
|
|
|
Db 84 GCGGAGGCGCACCGGGAGACCCCGGCTGTGTGTCGACGAGCGGATGTTGCGGCTG 140
|
|
|
```

```
QY 121 CCGGTGAGGACTTTCAGACCGATCTGTCGTCCCGGGGCCCGGACAGCCCGGAA 180
|
|
|
Db 141 CCGCTGAGAGACTTTCAGACCGATCTGTCGTCCCGGGGCCCGGACAGCCCGGAA 200
|
|
|
QY 181 CCGGACTTTCAGAAATGACCGCGGGGATACCGACAGCGGCTGCTCTGCGGCTC 240
|
|
|
Db 201 CCGGACTTTCAGAAATGACCGCGGGGATACCGACAGCGGCTGCTCTGCGGCTC 260
|
|
|
QY 241 TGCCTGCGGACACGAGGAGCATTCGCGGATCTTCGCGGAAACGCTGCGGCTCGCGGAA 300
|
|
|
Db 261 TGCCTGCGGACACGAGGAGCATTCGCGGATCTTCGCGGAAACGCTGCGGCTCGCGGAA 299
|
|
|
QY 301 CCGATGACAGCGCGGGATCTTCGAGGTGCGGACACCGGAGAGACGTTCTTCGCGGCTC 360
|
|
|
Db 300 CCGATGACAGCGCGGGATCTTCGAGGTGCGGACACCGGAGAGACGTTCTTCGCGGCTC 359
|
|
|
QY 361 CCGTGGCGGTTCAACGCGGCTGCGTACACTCTCTGAGCGGACCGACTCTCCGACGAG 420
|
|
|
Db 360 CCGTGGCGGTTCAACGCGGCTGCGTACACTCTCTGAGCGGACCGACTCTCCGACGAG 419
|
|
|
QY 421 CTGGAACCCCTGCGCTGAGGCGGACGAGGCTGTATGAGGCTGTGCGGACCGGAGAG 480
|
|
|
Db 420 CTGGAACCCCTGCGCTGAGGCGGACGAGGCTGTATGAGGCTGTGCGGACCGGAGAG 479
|
|
|
QY 481 CCGTGGGAGCGTCCAGTTCCACCGGAGTCCATCGGAGGACTTTCGCGGAGATC 540
|
|
|
Db 480 CCGTGGGAGCGTCCAGTTCCACCGGAGTCCATCGGAGGACTTTCGCGGAGATC 539
|
|
|
QY 541 ATGACCAATTCGCGGACCTGCGCTGCGTCCACACCGGAGCAACGCGGCGGACGAGGCGGAC 600
|
|
|
Db 540 ATGACCAATTCGCGGACCTGCGCTGCGTCCACACCGGAGCAACGCGGCGGACGAGGCGGAC 599
|
|
|
QY 601 TCCCGCTAAGAACTCCAGCTGCGCGGCTGCGAGCTGCTGCGGACCGGAGAGAGTACGC 660
|
|
|
Db 600 TCCCGCTAAGAACTCCAGCTGCGCGGCTGCGAGCTGCTGCGGACCGGAGAGAGTACGC 659
|
|
|
QY 661 CCGGCTGCGTGGCGGCGGAGGAGGACACGTTCTGAGTGAAGAGAGCTCCGCTCGAA 720
|
|
|
Db 660 CCGGCTGCGTGGCGGCGGAGGAGGACACGTTCTGAGTGAAGAGAGCTCCGCTCGAA 719
|
|
|
QY 721 GCGGCTGCGTGGCGTCTCTTCTGCGGAGGACCGCGGCGGCTGCGGAGTACCTCACC 780
|
|
|
Db 720 GCGGCTGCGTGGCGTCTCTTCTGCGGAGGACCGCGGCGGCTGCGGAGTACCTCACC 779
|
|
|
QY 781 TACCGGTGCGGAGCGGCGTCTCTGCGGCTGCGGAGGACCGGAGGACCGGAGG 840
|
|
|
Db 780 TACCGGTGCGGAGCGGCGTCTCTGCGGCTGCGGAGGACCGGAGGACCGGAGG 839
|
|
|
QY 841 CCGGCGGCGGCTTCTTCAACTACGAGGAGGAGGAGTGAACGCGGAGGAGGAGTCCCGTGGC 900
|
|
|
Db 840 CCGGCGGCGGCTTCTTCAACTACGAGGAGGAGGAGTGAACGCGGAGGAGGAGTCCCGTGGC 899
|
|
|
QY 901 CCGGAATGCTGCTTCAAGTTCAACTGCGGCTGCTGAGGCTTACCTGCGGCTTACGAGTGAAG 960
|
|
|
Db 900 CCGGAATGCTGCTTCAAGTTCAACTGCGGCTGCTGAGGCTTACCTGCGGCTTACGAGTGAAG 959
|
|
|
QY 961 GCGGAGACCAACGCGGAGACCCCGGAGCAGGCTCCCGGACCGGAGCGGCTTCTCTTC 1020
|
|
|
Db 960 GCGGAGACCAACGCGGAGACCCCGGAGCAGGCTCCCGGAGCGGCTTCTCTTC 1019
|
|
|
QY 1021 GCGGAGCGGAGCATGCGGCTTCAAGGACGAGGAGGAGTGTGCTTACCTGCGGCGGCTCGAG 1080
|
|
|
Db 1020 GCGGAGCGGAGCATGCGGCTTCAAGGACGAGGAGGAGTGTGCTTACCTGCGGCGGCTCGAG 1079
|
|
|
QY 1081 CCGCGGAGGACGAGCAGCGGCGGCGGCTGCTGAGTGTGCGGAGACGCGGACCGTCCACC 1140
|
|
|
Db 1080 CCGCGGAGGACGAGCAGCGGCGGCGGCTGCTGAGTGTGCGGAGACGCGGACCGTCCACC 1139
|
|
|
QY 1141 GCGCTGCGGCTGCGGCGGCGGCGGAGCGGAGCGGAGCGGCTTCTGCGGATCCCGAG 1200
|
|
|
Db 1140 GCGCTGCGGCTGCGGCGGCGGCGGAGCGGAGCGGAGCGGCTTCTGCGGATCCCGAG 1199
|
|
|
```



```

QY 1201 GCGGCGGCGGCTTCGCGCCCTGCGCCGCGCGCAAGACGCTTACCTGCAAG 1260
DB 1200 GCGGCGGCGGCTTCGCGCCCTGCGCTCGCGCAAGCGCAAGAGCGCT----- 1251
QY 1261 CGCATCAAGAGTCCCTCAAGAGATCCGCAAGCGCGAGTGTGAGATTTGCTTACC 1320
DB 1252 -----CGGCGCTCCGCAACGCGAGTGTGAGATTTGCTTACC 1292
QY 1321 AACATGTCAACCGCGCGAGACCGAGCGAGCGCGCTGCGCGCTTCTCGCGCTGCGCGCC 1380
DB 1293 AACATGTCAACCGCGCGAGACCGAGCGAGCGCGCTTCTCGCGCTTCTCGCGCTGCGCGCC 1352
QY 1381 ATGACCCCGCTCCGCTGAGCGCGCGCTGCTGAGTTCCCGCAAGTGTGCTGAGCGCC 1440
DB 1353 ATGACCCCGCTCCGCTGAGCGCGCGCTGCTGAGTTCCCGCAAGTGTGCTGAGCGCC 1412
QY 1441 TCGCGCGAGCGGTTCTTCAAGATGCGCGCGAGCGCGCGCTGAGTCCAAAGCCCATTAAG 1500
DB 1413 TCGCGCGAGCGGTTCTTCAAGATGCGCGCGAGCGCGCGCTGAGTCCAAAGCCCATTAAG 1472
QY 1501 GGGACCGCGCGCGCGCGAGACCGGAGAGAGAGAGCGCGCGCTGCGCGCGCGCGCGCGCGCC 1560
DB 1473 GGGACCGCGCGCGCGCGAGACCGGAGAGAGAGAGAGCGCGCGCTGCGCGCGCGCGCGCGCC 1532
QY 1561 GGGCGGAGAGAGAGACCGGCGCGAGAACCTGATGATGCTGACCTGTCCGCAAGACCTC 1620
DB 1533 GGGCGGAGAGAGAGACCGGCGCGAGAACCTGATGATGCTGACCTGTCCGCAAGACCTC 1592
QY 1621 AACAGCTCTGCGGATGCGCTCCGCTCAAGTGTCCCGCGCTTTGAGAGTGAAGACCTTAC 1680
DB 1593 AACAGCTCTGCGGATGCGCTCCGCTCAAGTGTCCCGCGCTTTGAGAGTGAAGACCTTAC 1652
QY 1681 GCGCGCGTGAACAGCTGTGTGTGACCATCCGCGAGAGCGCTGCGCGCGCGCGCGCGCGCGCC 1740
DB 1653 GCGCGCGTGAACAGCTGTGTGTGACCATCCGCGAGAGCGCTGCGCGCGCGCGCGCGCGCGCC 1712
QY 1741 GCGCGCTGCTAGCGCGCGCTTCCCGCGCGCTGCTGATGACCGCGCGCGCGCGCGCGCGCGCC 1800
DB 1713 GCGCGCTGCTAGCGCGCGCTTCCCGCGCGCTGCTGATGACCGCGCGCGCGCGCGCGCGCGCC 1772
QY 1801 ACCATGAGATCATTCGACCGCTGAGAGAGAGCGCGCGCGCTTCTGCGCGCGCGCTC 1860
DB 1773 CCCATGAGATCATTCGACCGCTGAGAGAGAGCGCGCGCGCGCTTCTGCGCGCGCGCTC 1832
QY 1861 GGAATGTTGCGCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 1920
DB 1833 GGAATGTTGCGCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 1892
QY 1921 GCGGACGCGCGCGAGTTCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1980
DB 1893 GCGGACGCGCGCGAGTTCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1952
QY 1981 GAGGAGTTCAACGAGACCGTGTAAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2040
DB 1953 GAGGAGTTCAACGAGACCGTGTAAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2012
QY 2041 GCGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2061
DB 2013 GCGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2033

```

# RESULT 4

```

US-10-819-386A-1
; Sequence 1, Application US/10819386A
; Publication No. US2005089884A1
; GENERAL INFORMATION:
; APPLICANT: Korea advanced Institute of Science and Technology
; APPLICANT: Shanghai Jiaotong University
; APPLICANT: LEE, Sang Yup
; APPLICANT: DENG, Zixin
; APPLICANT: CHEN, Shi
; APPLICANT: JEONG, Ki Jun
; APPLICANT: ZHOU, Xiufen

```

```

; TITLE OF INVENTION: Genes for the Synthesis of FR-008 Polypeptides
; FILE REFERENCE: P003-8015
; CURRENT APPLICATION NUMBER: US/10/819,386A
; CURRENT FILING DATE: 2004-04-06
; PRIOR APPLICATION NUMBER: KR10-2003-0074035
; PRIOR FILING DATE: 2003-10-23
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 138203
; TYPE: DNA
; ORGANISM: Streptomyces sp. FR-008
US-10-819-386A-1

Query Match      37.4%; Score 769.8; DB 21; Length 138203;
Best Local Similarity 63.2%; Pred. No. 8.2e-159;
Matches 1335; Conservative 0; Mismatches 682; Indels 96; Gaps 6;

QY 1 ATGCGACGGCTTCTGATTCGACACTGAGTCTGTTACCCGCAACCTGTTCCAGTATATC 60
DB 18610 ATGCGACCTTCTGCTGCACTGAGTCTGTTCACTTCACTTCACTTCACTTCACTTCACTT 18669
QY 61 GGGAGAGCCACCGGAGAACCCCGCTGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 117
DB 18670 TCCCGGAGCCACCGGAGAACCCCGAGTATCGGAGTATCGGAGTATCGGAGTATCGGAGTAT 18729
QY 118 CTGCGCGTGAAGACTTTCAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 177
DB 18730 GGTCTGCTGACCGCGTTCGACAACTGATGCTTCTCCCGGAGCGGAGTATCCCAACACCGC 18789
QY 178 GAAAGGAGCTTGGAAATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 237
DB 18790 CCGGCGAGCTTTCGCGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18849
QY 238 GTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 297
DB 18850 GTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18909
QY 298 GAAACCATGACCGCGCGGCTTCTGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 357
DB 18910 GAAACCATGACCGCGCGGCTTCTGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18969
QY 358 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 417
DB 18970 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19029
QY 418 GAGCTGGAACCCCTGCTGAGCGAGCGAGCGGAGTGTGATGAGGCTTGTGCGCGCGCGCGCG 477
DB 19030 GAGCTGGAACCCCTGCTGAGCGAGCGAGCGGAGTGTGATGAGGCTTGTGCGCGCGCGCGCG 19089
QY 478 AAGCGCTGTGAGGCGTTCAGTTTCAACCGGAGTTCATGCGAGCGACTTGTGCGCGCGCGAG 537
DB 19090 CTGCGCGCTGTGAGGCGTTCAGTTTCAACCGGAGTTCATGCGAGCGACTTGTGCGCGCGCGAG 19149
QY 538 ATCATGAGCGAACTTTCGCGCGAGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 591
DB 19150 ATCATGAGCGAACTTTCGCGCGAGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19209
QY 592 GGGGCGGACTCCCGCGTCAAGAACTTCAAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 651
DB 19210 GGGGCGGACTCCCGCGTCAAGAACTTCAAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 19269
QY 652 GAGGTACGCGCGCGCTGCTGCGCGCGAG----- 681
DB 19270 AGCGCACGCGCGCTCGGGTTCATTCGCGAGTGTGCGCGCGCGCGCGCGCGCGCGCGAGTTC 19329
QY 682 -----GGCACAGCTTGTGCTGAGCAAGAGCTTCCGTC 714
DB 19330 GCGTTCGACTGCTGTTTCGCAACCGGAGCAACCCCTTCTGTGCTGCAAGAGCGGTTC 19389
QY 715 CTCGAAGCGCGCTTCGCGCTTCTCTCTGCGCGAGAGCGCGCGCGCGCGCGCGAGTAC 774
DB 19390 GGGGCGGAGCTGGGCGAGCTTCTGCGATGGGCGAGCGCTTCAAGTTCCTTCCGCGCGAGCT 19449

```







Db 302 AGCCCCCAAGCCATCTGTCCACCGTCCGGACAAAGGCCAGACCTGTTTCAGAGGCC 361  
Qy 359 TCCCCCTGGCTTACCGCCCTGGCTTACCACTCCCTGGCCGCCACCGA---CCTCCCGC 415  
Db 362 TGCCCCAGACGATTACCGCGGTCCGCTACCACTCCCTGGTCCGTAACCGGAAACCGCTGGCAC 421  
Qy 416 ACAGACTGAAACCCCTGCTGAGACGACGAGGGTGTCTATGAGGCTGGCGGACCGC 475  
Db 422 TGAGCTTGAGACCCACCGCTGGGGGAGAGACGCGCTGTCTATGAGGCTGGCGGACCGCT 481  
Qy 476 AGAAGCGCTGTGGGGGCGTCCAGTTCCACCGAGTCCATCGGACGACCTTCGGCCGG 535  
Db 482 CCCGCGCGCTGTGGGGGCGTCCAGTTCCACCGAGTCCGTCCTACCGAGTTCCGGCCAC 541  
Qy 536 AGATCATGCGCAATTTCCGCGACCTCGCCCTCGCCACACCGGG-----580  
Db 542 GGATGTTCTGTGAACCTTCGCGAGCTGAACGCGCTCCCGCCCGGAGACTTCGACCGGCA 601  
Qy 581 -----580  
Db 602 AGGCGCGTACGTCGGAAACGCGACGACCGAGCCCGCTGCGGCCCTCCCAACG 661  
Qy 581 -----CACGCGCCACGCGGCGCACTCC 604  
Db 662 CCGAGGGGCGCGGGGAGACACCGCCGTCGTGTCCTCCCGCCCGCTGGTGGCCCGCT 721  
Qy 605 CGTACGAATTCACAGTGGCGCGCTGCACTGTGCGGACCGGAAAGAGTACCGCG 664  
Db 722 CCACCGCTTTCACACCGCGGAGATTCAGTGGCGCGTGGACGCGGAGCGGCTTCAACC 781  
Qy 665 GTCGCTGCGCGGAGGAGGACCACTTCTGCTGACAGACAGTCCGCTCCGAAAGCG 724  
Db 782 GGATGTACACGACGACGCGCCCGCGCTTCTGCTCGACAGTCCCGGATCGAGAGGAGAC 841  
Qy 725 CCTCGCGCTTCTCTCTCTCGCGACGACGCGCGCTCGCTCGCGAGTACCTCACTAC 784  
Db 842 AGTCCGCTTCTGCTTCTGCTGACGACACCGGACCGCTGGCGGAGTTGTAACGTAACG 901  
Qy 785 GCGTCCGCGACGCGTGTCTCGCTCCGCGCTTCGACGCGACCAACGACCGGACGCGC 844  
Db 902 ACATCGAGACCGGCTGTGAGATTCAGCGCGCCCGCGCGCTGTCGGAAGGTCAAGG 961  
Qy 845 GCCCTTCTTCACTACCTGAGAGAGAGCTGACGAGCGGAGCGCGGCTCCGCTGCGCCCG 904  
Db 962 CCAAGCTTCTGACTTACCTGAAACGACGATTTGAGACCGCTCAAGT---GATTCACG 1018  
Qy 905 AACTGCGCTTGAATTCACCTCGGCTACGTCGCTACCTCGGCTACGAGTGAAGCGG 964  
Db 1019 GACTGCGGCTTGAATTCACCGGCGGATATGAGGCTACTTTCGCTACGAGTGAAGCGG 1078  
Qy 965 AGACACCGGCGACCCCGGACACCGGTCCCGACACCCCGACGCGCGTCTCTTGC 1024  
Db 1079 ACTGCGGCTTCTGATCACTGATTCGAGGCGGAACTCCGACGCGCTGTGCTGTGCGCG 1138  
Qy 1025 ACCGCGCATGCGCTTGAACACCAAGAGAGGCTGCTGCTACGCTTGCACCGCC 1084  
Db 1139 ACCGCGTATGAGTGAACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1183  
Qy 1085 GGGGCGACGACGACGCGCGCGCGCTGCTGCGGAGAGAGCGCGAGAGAGAGAGAGAGAG 1144  
Db 1184 TCTGTCTCGCGGAGAACACCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1243  
Qy 1145 TGCGCTGCGCGCGCGCGCGAGCGAGCGAGCGCGCATGTCTTCGGAGTCCCGAGCGG 1204  
Db 1244 CGATGCGCGACGCTGATTTGTGTCTCGGCGCAAGCGAGCGCGCGCGCGCGAGCGG 1303  
Qy 1205 CGGCGCGCTTTCGGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1261  
Db 1304 CACCGCGCTTGAATGCGGTGAGCGCTGTGCTGCGCGAGCGCGTACCTTCGCGG 1363  
Qy 1262 GCATCGACGAGTGTCTCAAGAGATTCGCAACGCGCGAGTGTGAGAGTGTGCTGACCA 1321  
Db 1364 ACATCGGAGCTTCAAGCGGAACTCAAGAGCGGACCAATTCAGAGATTTGTCTGACCA 1423

Qy 1322 ACATGCTACCGGCGGACCGAGCGAGCGGCTTCTGCTTACTCCGCGCTGGCGCA 1381  
Db 1424 ACCGCGCAGAGCTGCGCGCGCGCTGATCGACCTTCAAGACTTCTACCGGATCTGTGTC 1483  
Qy 1382 TCAGCGCGCTTCCGTAACGCGCGCTGCTGAGTTCCCGAATCTGTGCTGAGCGCT 1441  
Db 1484 TCGACCGCGCGCGCTGAGCGCGCTTATCTGCGGTTCCGTAACCTGATGTGCGCGCTGT 1543  
Qy 1442 CGCGCGAGCGGCTTCTTCAAGTTCGCGCGCGCGCGCGCGCTGAGTCCAGCGCAATCAAG 1501  
Db 1544 CCGCGAGCGGCTTCTTCTGCGATCACCGCGAGCGCGCTGCGCGAGCGCGCGCGCGCG 1603  
Qy 1502 GAGCG 1561  
Db 1604 GACCG 1663  
Qy 1562 GCGGAGAGAGAGCGCGCGCGAGAACCTGATGATTCGAGCTGTGCGCGCAAGCTCA 1621  
Db 1664 CGAGCGACAGAGACCGCGCGCGAGAACCTGATGATTCGAGCTGTGCGCGCAAGCTCA 1723  
Qy 1622 ACAAGCTTGGCGGATTCGCTTCGTCACGTCGCGCGCTTTCAGAGTGAAGACTACG 1681  
Db 1724 GCGCGGCTTGGCGGACCGGAGCGGTGACCGCGCTGATGGGAGCGGAGCGGCTGCG 1783  
Qy 1682 GCGCGCTGACGAGCTGTGTGACCACTCGGCGGAGCGGCTGCGCGCGCGCGCGCG 1741  
Db 1784 CCACCGGCGACGAGTGTGTTCACCGCTGACGCGGCGCGCTGCGCGAGGCGATCGAGCGG 1843  
Qy 1742 CGCGCTGCGGAGCGCGCGCTTCCCGGCGCGCTCAATGACCGCGCGCGCGCGCGCG 1801  
Db 1844 TGAGCTGCGGATGCTGTGCTTCCCGAGGTTGCGGAGCGGCGCGCGCGCGCGCGCG 1903  
Qy 1802 CCATGAGATCATCGACCGCTGAGAGAGCGCGCGCGCTTCTTCCGCGCGCGCTG 1861  
Db 1904 CGATGAGATCATCGACTGCTGAGACGAGCGCGCGCGGCTGCTCGGAGCGCATCG 1963  
Qy 1862 GATGCTTCCCGCTGACGCGCGCGCGCGCGCTGACGATTCGATTCGCGACCATGTGCTG 1921  
Db 1964 GCTACTTCCGCTGACGAGTGGGCGCGGACCTCCGATTCGCGACCGCGCGCTGTCA 2023  
Qy 1922 CCGAGCGCGAGCGGAGGTTTCGCGTGGCGGCGGAGTTCGCTCCGCGAGAG 1981  
Db 2024 CCGAGCGGAGATGACCTCGCGCGCGGCGGCGGATTCGCTCGCTCGGATCGGCTG 2083  
Qy 1982 AGGAGTTCAACGAGACCGTGTAAAGCGCGCGCGCGCATGTGACCGCGCTGACGCGAGCG 2041  
Db 2084 GCGAGTACAGAGATGCTGTGAAGACGCGCGCGCATGCGCGCGCGCGCGAGCGCA 2143  
Qy 2042 CCG 2044  
Db 2144 TCG 2146  
  
RESULT 6  
US-10-156-761-1/c  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIRAMA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697



; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 1  
 ; LENGTH: 9025608  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces avermectilis  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (4187715)  
 ; OTHER INFORMATION: a, t, c, g, other or unknown  
 US-10-156-761-1

Query Match 29.4%; Score 606.2; DB 15; Length 9025608;  
 Best Local Similarity 58.3%; Pred. No. 1.8e-123;  
 Matches 1262; Conservative 0; Mismatches 763; Indels 138; Gaps 6;

QY 2 TGGGACGCTTTCGATGAGAACATGACATCTGTCACCCCAACCTGTTCACTATCATCG 61  
 DB 1470596 TGAAGACCTGCTCATGACAAATTAAGACTGTACAGTACAACTGTTCACACTGATCG 1470537  
 QY 62 GCGAGGCAACGCGGCAACCCCGCTGCTGTCGCCAAGACGCGCACTGTGCGCGCTCG 121  
 DB 1470536 CCGAGGTGAACGCGGAGGACCGGTGTGCTCTCATATGACGCCCGCGGAGACATTC 1470477  
 QY 122 CCG---TCGAGACTTTCAGACGATGCTGTGTCCCGGCGCCCGGACGCCGACCCGG 178  
 DB 1470476 CGGATCTCCGGGATTTTCGCAACGTGTGTGTGTGCGCGGCGCGGACACCCGCGAGA 1470417  
 QY 179 AAGGGACCTTCGAGATGACGCGCGCGGATTCACCAAGACGCGCTGCGCGCTCTCGGG 238  
 DB 1470416 GAGGTGACTTGTGATGTGGGAGAGTGTCTGCGCGCGCGCGCGCGCTGCTGTGGCG 1470357  
 QY 239 TCTGCTCGGCAACGAGGATGCGCCAGCTCTTCGCGGAAACGTCGCGCTCGCCCGG 298  
 DB 1470356 TCTGCTCGGCAACGAGGATGCGCGCGGAGAGCGGCGCTGAGTACCCCGGACCGG 1470297  
 QY 299 AACCCATGACGCGCGGCTTCGAGGTGCGGACACCGGCGAGAGAGTCTTCGCGGCGC 358  
 DB 1470296 AGCCCGCCACGCGCATCTGTCCACGTCGCGGCAACGCGGACGCTGTGAGAGGCG 1470237  
 QY 359 TCCCTCGCGCTTACCGCGCTGCTGCTACATCCCTGCGCGGACCGA---CCTCCCGC 415  
 DB 1470236 TGGCCGACGATTCACCGCGCTGCTGCTACATCCCTGCTGAGGAAACCGCTGCGAC 1470177  
 QY 416 ACGAGCTGCAACCCCTGCTGAGACGACGAGGCTCATGAGGCTTCGCGACCGCG 475  
 DB 1470176 TGAAGCTGAGACCGCGCTGCGGAGGAGACGCGGCTGATGAGATCGCGGACCGCT 1470117  
 QY 476 AGAAGCGCTGTGAGGCGTCCAGTTCACCCCGAGTTCATCGGACGACCTTCGCGCGG 535  
 DB 1470116 CCGGCGCGCTGTGAGGAGTCAAGTTCACCCCGAGTTCCTTCACGAGTTCGCGCAC 1470057  
 QY 536 AGATCATGCGCACTTCGCGGACCTGCGCTTCGCCCAACCGCG-----580  
 DB 1470056 GAGATTCGTAACCTTCGAGAGCTGACGCGCTTCGCCCGCGAGACTCGACGCGCA 1469997  
 QY 581 -----580  
 DB 1469996 AGGCGCTACGTCGGAACGCGACGACGACCGCCCGCTGCGGCGCCCGTCCCAAGCG 1469937  
 QY 581 -----CAGCGCGCACGCGGCGCGACTGCC 604  
 DB 1469936 CCGAGGAGGCGCGGAGCAGACCGCGCTGTGTCGCCCGCGCGCTGCGGTGCGCGCT 1469877  
 QY 605 CGTACGAATTCACGTCGCGCGCTGCACTGTGCTGCGCGGACGCGGAGAGTACGCGCG 664  
 DB 1469876 CCGACGCTTCACACCGCGGAGTTCGACGTCGCGGCGCGGAGCGCGCTTCACCC 1469817  
 QY 665 GCTGCTGCGCGGAGGAGACGATTCGCTGAGACGAGCTTCGCTCGAAGGCG 724  
 DB 1469816 GAGATGACACGAGCGCGCGCGCTTCGCTGAGACGAGCTCCCGGCTGAGAGAGGAG 1469757  
 QY 725 CCTCGCGCTTCTCTTCTCGGAGACGACCGCGCGCTCGCGGAGTACCTGACCTACC 784

DB 1469756 AGTCCGCTTTCGCTTCTTCGATGAGGACCGGACCGGCTGCGGCTTCGATCGGTACG 1469697  
 QY 785 GCGTCCGCGACGCGGCTGCTCCTCGCTCGCGGCTCCGACGCGACACGACCGCGACCGCGC 844  
 DB 1469696 AGCTCGAGAGCGGCTCTGTGAGATGAGCGGCGCGCGCGCGCGCGCGGAGTACGAG 1469637  
 QY 845 GCGCTTCTTCAATCACTGAGAGAGACGCTGGAACGCGGAGTCCCGTCCGCGCGCG 904  
 DB 1469636 CCAAGCTTTCGATCTGAGAAACGCGCACTGTGTGACCCGTACAGT---GGAATGCCACG 1469580  
 QY 905 AACTGCTTCGAGTTCACCTTCGCTACGTCGCTACCTCGGCTTCGAGCTGAGGCGG 964  
 DB 1469579 GACTGCGCTTCGACTTCACCGCGGAGTATGTGGGCTACTTTCGCTACGAGTGAAGCGCG 1469520  
 QY 965 AGACACGCGGCGACCGCGCGACCGGTCCCGGACCGCGCGCGCGCTTCGCGCG 1024  
 DB 1469519 ACTGCGGTTCCGTAACCGTCACTGAGGCGGAACTCCGACGCTGCTGCTGCTGCGCG 1469460  
 QY 1025 ACCGCGCATGCGCTTCGACCAACGAGAGGCTGTGCTTACCTGTGCGCTTCGACCGCG 1084  
 DB 1469459 ACCGCGTATTCGAGTGAACATCAAGAGCGA-----GCCACTTACGCG 1469415  
 QY 1085 GGGGCAACGACGAGCGCGCGCGCTGCTGCGGAGACGCGCGACCTTCACCGCGC 1144  
 DB 1469414 TCTGTCTCGCGAGAACACCGCGAGGCGGCAAGGACCGCGCGACTGCTTCGACGCG 1469355  
 QY 1145 TGGCGCTGCGCGCGCGCGCGGCGGACCGCGCGCATGTGCTTCGAGGATCCCGAGCGCG 1204  
 DB 1469354 CGATGCGCACTCACTTCTGTCTGCTGCTGCGGCGGCGCGCGCGCGCGCGCGCG 1469295  
 QY 1205 CGGCGCGCTTTCG 1261  
 DB 1469294 CACCGCGCTTCGATGCGCGGAGCGGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCG 1469235  
 QY 1262 GCATGACGAGTGTCTCAAGAGATCCGACGCGGAGTCTGACGAGTCTGCTTACCA 1321  
 DB 1469234 ACATCGGAGCTTCGCAAGCGGAACTCAAGAGCGGCGCACGATTACGAGTCTGCTGACCA 1469175  
 QY 1322 ACATGCTACCGGCGCGGACGAGGCGCGCGCGCTGCGCTTACTCGCGGCTCGCGCGCA 1381  
 DB 1469174 AGCGGCGAGCTGCG 1469115  
 QY 1382 TCAGCCCGCTCCGTAACGCGCGCGCTGCTGAGTTCGCCGAATGTCGTCGAGCGCT 1441  
 DB 1469114 TCGACCG 1469055  
 QY 1442 CGCGCGAGCGGTTCTCAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1501  
 DB 1469054 CCGCGGAGGTTTTCGCGGATTCACCGCGAGCGCGCTCGCGAGCGCAACCGCGTCAAG 1468995  
 QY 1502 GAGCG 1561  
 DB 1468994 GACCG 1468935  
 QY 1562 GCGGAGAGAGGACCG 1621  
 DB 1468934 CGAGCGACAAAGACCG 1468875  
 QY 1622 ACAGCGTTCGCGGATCGCTCCGTCACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1681  
 DB 1468874 GCGGCGTTCGCGGACCGCGGAGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCG 1468815  
 QY 1682 CGCGCGTTCGACGAGTGTGACCATCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1741  
 DB 1468814 CCAAGTTCGACCAAGTGTGCTCCACGTCGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1468755  
 QY 1742 CGCGCTGCTACGCGCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1801  
 DB 1468754 TGAAGCTGCTACGTCGCTTCGCGGAGTTCGCTGACCGGCGCGGAAACGTCGCA 1468695  
 QY 1802 CCAATGAGATCAATGACCGCGCTGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1861



Db 1468694 CGATGAGATCATGACTCGCTGAGAGACGAGCCCGGGGTGTACTCCGAGCCATCG 1468635  
 Qy 1862 GATGTTGCGCCTGAGCGGGCGCGACCTCAGCATGTGTATCCGACCATGCTGTGG 1921  
 Db 1468634 GCTACTTGGGTGAGTGGGGCGCGACCTCGCCATGCGACGCGCGTGTCA 1468575  
 Qy 1922 CCGACGGCCGAGCGGAGTTGCGCTCGCGGGGCGATGTGCTCTCCGACGAGAG 1981  
 Db 1468574 CCGACGGGAGATGTCACCTCGGCGCGGGCGGTGGATGTCTCGCTCCGATCGGTG 1468515  
 Qy 1982 AGAGTTACCGAGACCGTGTAAAGCCCGCGCATGTATCACCGCTTCGACGCGAGC 2041  
 Db 1468514 GCGAGTACGACGAGATGCTGTGAAGACGCGCGCATGTGCGGCCACCGAGCCGGA 1468455  
 Qy 2042 CCG 2044  
 Db 1468454 TCG 1468452

RESULT 7  
 US-10-156-761-1  
 ; Sequence 1, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; PRIOR FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 1  
 ; LENGTH: 9025608  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces avermitilis  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (4187715)  
 ; OTHER INFORMATION: a, t, c, g, other or unknown  
 US-10-156-761-1

Query Match 14.6%; Score 301.4; DB 15; Length 9025608;  
 Best Local Similarity 49.7%; Pred. No. 4,4e-57;  
 Matches 948; Conservative 0; Mismatches 936; Indels 25; Gaps 6;  
 Qy 135 CGACGGAGTGTGTGTCCTCCCGGGCCCGGACGCCCCGAGACCGGGAGACGGAGATTCTCGGAAT 194  
 Db 7417356 CGGCAAAATGTGTGCGATGACGACAGCCCGACGCGGAGGCGGACGCTGTGACCG 7417415  
 Qy 195 CAGCGCCCGGGCGATGACGACGAGCGGCTGCTCGGCGTGTGCTGCTCGGCGGACCA 254  
 Db 7417416 CGAGATCGCGCGCGCTCAAGGAGGAGCGCGCGGCTGTGCTCCCGCATTCGCCAGCA 7417475  
 Qy 255 GGGGATCGCCAGCTCTTGGCGGAAACGTTGGCTTGGCCCCGGAAACCATGACGAGCGG 314  
 Db 7417476 GTACGACACCGGTGAGTGTGTCTCGCTGATGATGAGAGAGCGCTGACCGGAC 7417535  
 Qy 315 GGTCTCGAGGTGCGGACACGCGGAGAGAGCTTTCGCGGGCTCTCCCTCGCGCTTCA 374  
 Db 7417536 GCTTACACGAGCGCGCTGACCTACTGTGTGCGGAGCGGGGAGTACTGGTGAAGG 7417595  
 Qy 375 CGCGGTGCGTACACTCTCTGCGCGACCGACTTCCCGAGAGCTCGAACCCCTCGC 434  
 Db 7417596 CGACACCTCGCGGACGTCAGTGAAGTCCGTAGGCTTGACTGGAGCCCGACAC 7417655

Qy 435 CTGAGAGACGAGGGGTGTCTATGGGCTTCCGCAACCGGAGAACCGCTGTGGGCGCT 494  
 Db 7417656 GTGTCTGTAAAGTGTGACCAAGGTGGCGCGCTTCCACACGCGGCGGACAGCTGTCT 7417715  
 Qy 495 CCAATTTCACCCGAGATTCATCGGACGCACTTGGCGGGAGATCATGGCAACTTCCG 554  
 Db 7417716 CAGCGCGAGTGTCTCTAAGAGCGGCAATTCGGCGGTATCCCGACGGATATAGTAAG 7417775  
 Qy 555 CAGCTCGGCTTGGCCACACCGGACGCGGCGCAAGGAGCGGAGCGGAGTCCCGTCAACT 614  
 Db 7417776 TCA----GCCGACATGAGCTGTGAGAGCTTCCGCAAGTGGCCACGAGCGGCTGAT 7417831  
 Qy 615 CCACGTGCGCGGTGTGAGTGTGCTGCGGACCGGAGAGGTATGCGCGGCTGCTGCC 674  
 Db 7417832 TCCGTCACGCGCAAACTCTGCGCGACGAGCAACCCGCTGCGGCTTACCGCAAGCT 7417891  
 Qy 675 CGCGAGAGGACACGCTTGTGCTGAGACAGCTGCTGCTGAGAGGCGCTGCGGCTT 724  
 Db 7417892 CGCGCGGAGCGCGCGGACCTTCTCTGAGATCGCGAGAAAGCGCGCTTGGCGCTT 7417951  
 Qy 735 CTCCTTCTGCGGAGACCGCGGCGCTTCCGAGTACTCACTACCGCTGCGCGA 794  
 Db 7417952 CAATGATCCCGTACTCTCTGTGGGCGTCCGCTCGCGGACCTCACGAGGCTGA 7418011  
 Qy 795 CGGCTGTCTCTGCTGCGGCTTCCGAGCACAGACCCGAGCGCGGCGCTTCTT 854  
 Db 7418012 CGGCGAGGCGCACTGTGCTGCGACCCCGCGTGGGCGTCCCGACGAGCGGACCGCT 7418071  
 Qy 855 CAATCACTGAGAGACGACTGAAAGCGGAGCGGCTGCGGCGGCGGCGGAGTGCCTT 914  
 Db 7418072 CGCGCGCTTGGCGGCGGACCATGAGGCGCTGACACGCGCCACGAGAGGCGCTTGC 7418131  
 Qy 915 CGAGTTCAACTCGGACTAGTGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 968  
 Db 7418132 C--TTACCGGCGGAGTGTGGGTATCTGGCTTACGACATGTGCGCGCTGAGAA 7418188  
 Qy 969 CACCGGAGACCGCGGACCGGCTCCCGGACCGGCGGCGGCTTCTTGGCGGACCG 1028  
 Db 7418189 GATCGGCGCGGCGGAGCGGAGTGTGAACCTTCCGAGTGAACATCTCTTCAACAG 7418248  
 Qy 1029 CGCATGCGCTGACACACAGAGAGGCTGTCTGCTGCGGCTGCGGCGGCGG 1088  
 Db 7418249 CGATCTGCGCTGATGAGACATGAGGAGGCTGCTGCTGATGCTCAAGCGATCAA 7418308  
 Qy 1089 CCACGACGAGCGGCGCGGCTGCTGCTGCGGAGAGCGCGGACCTTACCGGCTGCGC 1148  
 Db 7418309 CCACACGACCTGACAGCGGCGTGCAGAGGCTTACGCGGAGAGCGGCTTCCGA 7418368  
 Qy 1149 CGTCCGCGCGCGCGGAGCGGACCGCGGCTGTGCTTGTGAGATCCCGAGCGGCGG 1208  
 Db 7418369 CCGCATGAGGCGGACCTGTGCGCGC-----GGTCCGAGACCGCGCGCGCTTCC 7418422  
 Qy 1209 CGGCTTGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1268  
 Db 7418423 GCGCTTCAAACTCCGAGTACACGCGCTGTGGGCGGCGGCGGCGGCGGCGGCGG 7418482  
 Qy 1269 CGAGTGTCTCAAGAG--ATCCGCAACCGGAGTGTGATGAGATGTGCTGACCAACT 1325  
 Db 7418483 CAGAGACATCAAGAGCGGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7418542  
 Qy 1326 GGTCAACCGCGGAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1385  
 Db 7418543 CTTTGAAAGCGCGGTGACGCGGCAAGCGGCTGTGACCTTACCGGATATGCGGCGACCA 7418602  
 Qy 1386 CCGCGTCCGATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1445  
 Db 7418603 CCGCTTCACTGATGATGATCTTTCGCTTGAAGGCTTCAAGTGTGAGTGTGCTGCTCC 7418662  
 Qy 1446 CGAGCGGTTCTCAAGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1505  
 Db 7418663 CGAGGCGCTGCTGAG--GTCAGAGAGCGGCGGAGGCGATGTGTCAACCCATGCGGCGAC 7418719







QY 1703 CGACCATCCGGGGAAGGCTGCGGCGCCGACACGACCGCCGCTGCGTACGCGCGCT 1762  
D 1133 CGACCGTGAACCGGCGGGGTGCGGCGGGGCGCTACGCGCTTGAAGTGTGACGCGCTGCT 1192  
QY 1763 TCCCCGGGCGCTCCATGACCGGCGCGCCGAAAGACCGACCATGAGATCATCGACCGCC 1822  
D 1193 TCCCGGCGCGGACCGCTCTCCGCGCGCCGCAAGCCCGCGCCCTGCAATCATTCAGCAGAC 1252  
QY 1823 TGGAGGAAGGCGCGCGCGCTCTACTCCGCGGCGCTCGATGATGCTGCGCTTCAAGCGCG 1882  
D 1253 TGGAGCGCTCCGCGCGGCGCTGACGCGGCGCTGCGCTGCTACCTTCCGCGCG 1312  
QY 1883 CCGCGCATCTGAGATGCTCATCCGACCATGCTGCGCGACGCGCGCGAGTTGCG 1942  
D 1313 ACTCGACGACCGCATGCGCATCCGACGCGGCTGCTGCGGACGCGACGCGCTACGTCG 1372  
QY 1943 GCGTCGCGCGGCGCATGCTGCTGCTTCCGACGAGAGAGAGTTGACCGGACGCGTGG 2002  
D 1373 AGCGCGCGCGCGCATGCTGCTGCGGACCGCGCTGCGCGAGAGAGAGAGTTGCGCA 1432  
QY 2003 TAAAGCGCGCGCATGCTGCTGCTGCGCGCTGCGAC 2034  
D 1433 ACAAGCGCGCGCGCTGCTGCGCGCGCTGCGAC 1464

RESULT 9  
US-10-716-803-9  
; Sequence 9, Application US/10716803  
; Publication No. US20040229236A1  
; GENERAL INFORMATION:  
; APPLICANT: Blanc, Veronique  
; Crouzet, Joel  
; Jacques, Nathalie  
; Lacroix, Patricia  
; Thibaut, Denis  
; Zagorec, Monique  
; Debussche, Laurent  
; De Crecy-Lagard, Valerie  
; TITLE OF INVENTION: Polypeptides Involved in The  
; Biosynthesis Of Streptogramins, Nucleotide Sequences  
; Coding For These Polypeptides And Their Use  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flunegan, Henderson, Parabow, Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/716,803  
; FILING DATE: 20-Nov-2003  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 09/635,359  
; FILING DATE: 09-AUG-2000  
; APPLICATION NUMBER: US 09/231,818  
; FILING DATE: 15-JAN-1999  
; APPLICATION NUMBER: US 08/403,852  
; FILING DATE: 10-MAY-1995  
; APPLICATION NUMBER: PCT/FR 93/00923  
; FILING DATE: 25-SEP-1993  
; APPLICATION NUMBER: FR 92/11441  
; FILING DATE: 25-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arisgo, Salvatore J.  
; REGISTRATION NUMBER: 46,063

REFERENCE/DOCKET NUMBER: 03806;0054-04000  
TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 408-4000  
; TELEFAX: (202) 408-4400  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 645 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: S.priestinae spiralis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 61..645  
; OTHER INFORMATION: /product= "gene papa"  
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:  
US-10-716-803-9  
Query Match 13.2%; Score 272.2; DB 20; Length 645;  
Best Local Similarity 68.7%; Pred. No. 4.4e-50;  
Matches 392; Conservative 0; Mismatches 173; Indels 6; Gaps 1;  
QY 2 TGGCGACGCTTGTGATGACAACTAGACTGTTACCCCAAACTGTTCAATGATCG 61  
D 71 TGGCAACCTGCTGATGACAACTAGACTGTTCACTTCAAACTTTCAGATGCTGG 130  
QY 62 GCGAGGCAACCGGCGCAACCCCGCTGCTGCTGCGCAAGACCGCGCATGCTGCGGCTCG 121  
D 131 CCGAGGTGAACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 190  
QY 122 CCGTCGAG-----GACTTGAAGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 175  
D 191 CCGTCGAGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 250  
QY 176 GGGACGGAATTGGGAATGACCGCGCGGCGGATGACCGGACCGGACCGGCTGCGCTGCTG 235  
D 251 CCGACACCGAAGCTGCGGCTGACGCGCGGCGGATGACCGAATGAGACTGCGGCTGCTG 310  
QY 236 GCGTCGCTGCGGCGCAACGAGGATGCGGCGGCTGCTGCGGCGGAACTGCGGCTGCTG 295  
D 311 GCGTCGCTGCGGCGCAACGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 370  
QY 296 CCGAACCTGACGAGCGCGGCTGCTGCGAGGTGCGGACGACCGGCGGAGCGTCTTCCGG 355  
D 371 CCGAACCTTTCACGCGCGCGGCTGCGGACGATGCGGACGAGGCGCTGTTGCGA 430  
QY 356 GCGTCGCTGCGGCGTTCACCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 415  
D 431 ACATCCCTCCCGCGGACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 490  
QY 416 AGGAGTCAACCCCTGCTGCGTGAAGGACGAGAGGCGGTCTATGCGGCTTGGCGACCGG 475  
D 491 CCGAGCTGCGCGCACCGCGCGCACCGCGCGAGCGGAGCTGATGCGGCTGCGCGCACCG 550  
QY 476 AGAAGCGCGTGGGCGCTGCGGCTGCGGAGTCCACCGGAGTCCATGCGGAGCGGCTTGGCGG 535  
D 551 ACCTGCGCGCTTCCGCGGCTGCGGAGTCCACCGGAGTCCATGCGGAGCGGCTTGGCGG 610  
QY 536 AGATCATGCGCAACTTCCGCGACTGCGGCT 566  
D 611 GGATGCTGCGCAACTTCCGCGACTGCTGCT 641

RESULT 10  
US-10-282-122A-30135  
; Sequence 30135, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangu



```

/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlson, Kari
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trivick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: EUTRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282.122A
/ PRIOR FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 30135
/ LENGTH: 1362
/ TYPE: DNA
/ ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30135

Query Match      13.1%; Score 270; DB 17; Length 1362;
Best Local Similarity 58.7%; Pred. No. 1.2e-49;
Matches 487; Conservative 0; Mismatches 340; Indels 3; Gaps 1;

QY      1227 CCGCGCGCGCCGACGACGAGAGCGCTTACTCAAGCGCATGACGAGTGCCTCAAGAGAT 1286
DB      528 CCGACGAGCATGAGCGCATCCGACTATCGGAGGCGATCCGCGCATCCAGAGACTACAT 587

QY      1287 CCGCAACGGGAGTTCGACAGATCTGCTCGACCAACATGATGACCGGCCGACCGAGC 1346
DB      588 CCAAGGAGGAGTCTGACGAGTGAACATATACCGCGCTTCCAGCGCCGCTTGACGCG 647

QY      1347 GACGCGCTTCCGCTTACTCCGCGCTGCGCGCATGAGCCGTCGCGGACCGCCCT 1406
DB      648 CTGCGCGTGGCGCGGCTATGCGCGCTGCGCGGAGCGCTGCGCGCATCGCGCTTCTCCGCTA 707

QY      1407 GCTCGAGTTCCCGAACTGCTGCTGAGCGGCTGCGCGGAGCGGTTCTCAAGATCG 1466
DB      708 CTGCGACTGCGCGAGCGCGCATCTCACTGTGCGCGGAGCGCTTCTCAAGCTCGG 767

QY      1467 CGCGGAGGCGGCTGAGTTCAGGCTCAAGCCATCAAGGGGACCCGCGCGGCGGACCGC 1526
DB      768 ---CAAGGGGCAAGTGGAAACCGCGCGATCAAGGGGACCCGCGCGGCAAGACCCC 824

QY      1527 GAGGAGGAGGAGCGAGCTCCGCGCGCGACCTGCGCGGCGGAGAGAGGACCGGCGGAGAA 1586
DB      825 CGAGGAGGAGATGCGCTGCGCGCGCTGCTGCTGCGCGGAGAGGACCGCGGAGAA 884

QY      1587 CCGATGATGTCGACCTGTGTCGCAACGACTCAACAGCGTCTGCGCATCGGCTCCG 1646
```

```

DB      885 CCGATGATGTCGACCTGCTGCGCAACGACATCGGACCGCACTTCCAACTTGCGAGCGT 944
QY      1647 CCACGTGCCCCGCGCTTTGAGGTGAGACCTACGCGCCCGTGGACACCACTGGTGTGAC 1706
DB      945 ACCGTTACCGAGACTGTGTTGCCCTGGAAAGCTTATCCAACTGATCACCCTGGTAGAG 1004
QY      1707 CATCCGGGAGCGGCTGCGCGCGGACAGCACCGCGCTCGTACGCGCGCTTCCC 1766
DB      1005 CCGTCAACGGCGAACTGCGCGCGGCGGAGAGACCGCTTCGACCTGCTGGAAGGACGCTTCCC 1064
QY      1767 CGCGGCTTCATGACCGCGCGCGCCCAAGAGCGCACCATGAGATCATGACCGCTGGA 1826
DB      1065 CGCGGCTGATCATCAGCGCGCGCGCAAGATTGCGCGCATGAGATCATGACGAGCTGGA 1124
QY      1827 GGAAGGCCCCCGGGGCTTACTCCGCGGCGCTCGGATGTTGCGCCTCAGGCGCGCCC 1886
DB      1125 ACCGAGCCGACCGGGGATCTACTGCGGACGCTGTTTACTCTGAGTGTGCGCGAGAT 1184
QY      1887 CGACTCAGCATGCTCATCCGACCATCGTGTGCGCGGACGCGCGAGTTCGCGCT 1946
DB      1185 GGAACACTGATGTGCATCGCACCTGTGTGACAGGACGCGCATGATGCTGCGGG 1244
QY      1947 CGCGGCGGAGATGCTGTCTCTTCCGACGAGAGAGATTACCGAGACCGTGTAA 2006
DB      1245 CGCGCGCGCATGCTGCGCGCATCGCATCGGAGGACGAGTACAGAGAAACCTGAGCAA 1304
QY      2007 GCGCGCGGCGCATGTGACCGCGCTGAGCGAGCGCGCGTGGGCGGCGCC 2056
DB      1305 GGTCCGGGTGCTGTGGAAACCTTGAAAGATTGCCGGGAGACGCTCC 1354
```

## RESULT 11

```

US-10-501-282-17
/ Sequence 17, Application US/10501282
/ Publication No. US20050203280A1
/ GENERAL INFORMATION:
/ APPLICANT: MCWICHAEL, JOHN CALHOUN
/ APPLICANT: ZAGURSKY, ROBERT JOHN
/ APPLICANT: RUSSELL, DAVID PARRISH
/ APPLICANT: FLETCHER, LEAH DIANE
/ TITLE OF INVENTION: ALLOIOCCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
/ FILE REFERENCE: AM100780.12
/ CURRENT APPLICATION NUMBER: US/10/501,282
/ PRIOR FILING DATE: 2004-07-09
/ PRIOR APPLICATION NUMBER: 60/333,777
/ PRIOR FILING DATE: 2001-11-29
/ PRIOR APPLICATION NUMBER: 60/426,742
/ PRIOR FILING DATE: 2002-11-18
/ PRIOR APPLICATION NUMBER: PCT/US02/36123
/ PRIOR FILING DATE: 2002-11-25
/ NUMBER OF SEQ ID NOS: 6653
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 17
/ LENGTH: 2046
/ TYPE: DNA
/ ORGANISM: Alloiococcus otitidis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (22) ..(2043)
US-10-501-282-17
```

```

Query Match      13.0%; Score 267.2; DB 22; Length 2046;
Best Local Similarity 48.8%; Pred. No. 4.5e-49;
Matches 999; Conservative 0; Mismatches 963; Indels 84; Gaps 7;
```

```

QY      1 ATGCGAGCGCTCTGATGCAACAAGTACGCTGTACCCACAACTGTTCCAGTATC 60
DB      22 ATGAAGTCACTATTCATATCACTAATGATTTTATTACTTAACTTATACAGCTTAT 81
QY      61 GCGGAGGCGCACCGGCAACCCCGTGTGTCGCCCAAGAGCGCGACTGTGCGGCGT 120
```



Db 82 GGGAGGTCAACAGGAGGAGCCATGCTGTAATAAATGACAGATGACTTACCAAGAA 141  
QY 121 CCCGTCGA---GACCTTGAAGCGATGCTGTCTCCCGGAGCCCGGACGCGGACCGG 177  
Db 142 CTATTGATGTCGACTTTGATTAATGCTATTATTTACACCGGGCCCGGATGTCGGAACGG 201  
QY 178 GAACGGGACTTTCGGAATCAGCCGCGGCGATCAACGACGGGCTTCCGCTTCGCGC 237  
Db 202 GACAAAGACTTCGCGCTTGTCCGCGCAAGTAATGAAAGCTTGAACAAGCTTATCTTTGGC 261  
QY 238 GTCGCGCTCGGCGCAACAGGCGATGCCAGCTCTTCCGCGGAACCGTCCGCTCGCCCG 297  
Db 262 ATTTGCTGCGGCAACAGGCGATTTATTTATTAACCGTGCACCTTAATGCGGCGAT 321  
QY 298 GAACCCATGCAACGCGCGGCTCTCCGAGTGCGGACAACCGGCGAGAGCTCTTCCGCGG 357  
Db 322 ATCCCATGCAATGCTCGGCAAGATTCGCTTACCATTAATGGAAAAATCTTTTGAAGCG 381  
QY 358 CTCCTCGCGCTTACCGCGCTGCGTACCACTCCCTGG---CCGCGACCGACTTCCCG 414  
Db 382 ATAGACCAAGGCTTTGAAGTCAACCGCTACCACTCTGTTGTGAAGCAAGAAATTA 441  
QY 415 GAGAGCTCGAACCCCTCGCTGAGAGGAGCGAGCGGCTGTCATGGCTTGGCGACCGC 474  
Db 442 GAGAAATTTCAAAATCGATGCGCAAAACCGATGGAATTTGTCATGCGCTGTCCCAAG 501  
QY 475 GAGAGCGCGCTGCGGCGTCCAGTTCCACCCGAGTCACTCGGACGAGCTTCGCGCG 534  
Db 502 ACCAAGGCCATCTATGGGGTCAATTCACCCGAAATTCATGTCACCGATGAGGAAA 561  
QY 535 GAGATCATGCGCACTTCGCGACCTCGCCCTCGCCACCAACGCGGACGCGCGCAAGG 594  
Db 562 AAATGATGAAAACTT-----TATGCTTGTGCGAGAC 597  
QY 595 GCGGACTCCCCGTCGAACTTCACGTGCGCGCGCTGAGCTGCTGCGGACGCGGAAG 654  
Db 598 TACTATTAACCAATTCAGCTTTTATTAATAAAGTTCCGGAACTAGGCAACGCAAG 657  
QY 655 GTACGCGCGGCTGCGCTGCGCGGAGGAGGACACGCTTGGCTGGAACAGCTCCGTC 714  
Db 658 CTCTACGATCAATTTGGCCAAATTTGATGACCAAGTACTTTGGCTGCTCACAGAGTG 717  
QY 715 CTCGAAAGCGCTTCGCGCTTCTCTTCTCGCGGACGACGCGGCGCGCTCGCGAGTAC 774  
Db 718 GAGCGTGGCTTGTCAAGGTTTTTCATCTTTGCGATGCTGACCAAAAGCGGCGCACACC 777  
QY 775 CTCACCTACCGCGCTGCGGAGCGGCTGCTCCGTCGCGCTCCGACGCGCACGACCC 834  
Db 778 TTAATAATACGATGTCGACCAAAAGAAAGTGAATAAAGATAGCCAGTCCGACAGGT 837  
QY 835 CGGACGCGGCGCTTCTTCAACTACCTGAGAGAGAGCTCGAACGCGAGCGGCTCCC 894  
Db 838 GAAA---GCTTTCAGACTGATCTTTTCTTAATAAACCAACGACCAAAATGCGCT 894  
QY 895 GTCGCGCCCGGACCTGCGCTTGAAGTCACTCGGCTAGTGGCTACTACCTCGCTACGAG 954  
Db 895 TATGAGCGCGCTTGGCTTTTGACTTTCACTAGGCTATATGCTATATGCTGCTAGAG 954  
QY 955 CTGAAAGCGGAGAC---CAACGCGGAGCCCGGACCGGCTCCCGGACCGCGAGCGCG 1011  
Db 955 GTCAAAAAAGATCAAGTTCAAGTTCACCAACAGACAGTCAACCTATCCGAGTGGCTAC 1014  
QY 1012 TTCTCTTGGCGAGCGCGCTGCTGCTGACCAACGAGAGGCTGCTGCTACTGCTG 1071  
Db 1015 TTACTATTATGATACCGGCGCTTGTGCTATGACCAACGAGGAAAGGACTTTTACTTCTC 1074  
QY 1072 GCGCTGACCGCGGCGGCGACAGACGCGCGCGCTGCGGAGAGAGCGCGGAG 1131  
Db 1075 TCCCTACCA-----GATGAC 1089  
QY 1132 ACCCTCAACCGGCTGCGCGCTGCGCGCGCGGCGAGCGGACCGCGCATGCTTTCGCG 1191  
Db 1090 CAGGATTGATTCAGGCGTCAAGAGCAGGCTCAGAGGCGCTCAAAAGTTTCAGCAAGAA 1149

QY 1192 ATCCCGAGCGCGCGCGCTTCCGCGCTGCGCGCTGCGCGCGCGGCGGACGAGCGCC 1251  
Db 1150 GAGACCAAGAGGCAAGCAAGCTTCCACAGCTGCGCTTGG---TCAAGCAAAAGCAGCC 1206  
QY 1252 TACCTCAAGGCGATGCAAGTGGCTCAAGAGATTCGCAACGCGGAGTGTACGAGATC 1311  
Db 1207 TATATCAAGACATTTGAATAACCATTCGAAGCTGATCAAGGCTGGGAAAGTTATGAAGTT 1266  
QY 1312 TGGCTAACCAATGCTACCGCGCGGACCGAGGCGACCGCTTCCGCTTACTCTCGCG 1371  
Db 1267 TGGTTGACCAACCGGCTGATATTTGAGGCTCAATTTGAGCTTCCCTTATTAACAACAC 1326  
QY 1372 CTGCGCGCATTCAGCCCGCTCCGTAAGGCGCGCTGCTGAGTTCCTCCGACCTGCGGTG 1431  
Db 1327 TTGCGCGAGAAAGTCCCGCTCAGTACTCTGCTTTTGGCTTTAATGATTTTCAAGTT 1386  
QY 1432 CTGAGCGCTCGCGCGGAGGCTTCTCAAGATGCGCGCGGACGCGCGCTGAGTCCAG 1491  
Db 1387 CTTCCTCTTCCATGAGGCTTTTATTAATGTCGACAAAGGACCGGCTTGTGACACTAG 1446  
QY 1492 CCCATCAAGGAGACCGCGCGCGCGCGCGCGCGCGAGAGGACGAGCGGCTCCGCGCC 1551  
Db 1447 CCCATCAAGGAGGAGCGGCTCAGACGCGGAGACGATGACCAAGAGACCAAGCTTATGA 1506  
QY 1552 GACCTGCGCGCGCGGAGAGGACCGGCGCGAGAACCTGATGATGCTGACCTGCTCCG 1611  
Db 1507 GGCCTCGCGAGTGAAGAAACCAAGGACGAGAAACCTGATGATGCTGACCTTTCGCG 1566  
QY 1612 AACGACTCAACAGCTCTGCGGATCGGCTCCGTCACGTCGCGCGGCTCTTTCGAGTG 1671  
Db 1567 AATGATCTGGCGCGCTTTTGTGAAATCGGCTGCGTTGAGGTACCCAAATGATGATGTA 1626  
QY 1672 GAGACTTACGCGCGCGCTGCAACAGCTGTGTGACATCCGAGGACGCTGCGCGCGCG 1731  
Db 1627 GAGACTTACGCACTCCACAGGCTGATGACCAAGTTCACGAGTTCAGGCGGCTCAAGAA 1686  
QY 1732 ACCAGACCGCGCGCTGCTGACCGCGCGCTTCCCGGCGGCTTCCATGACCGCGCGCC 1791  
Db 1687 CTAGATGTGTGAAGTGTGAAGAAACACTTCTCCGCGGCTGCAAGTGGGCGCT 1746  
QY 1792 AAGAACGCGACCATGAGATCATTCGACCGCTGAGAGAAAGGCGCGGCGGCTCTACTCC 1851  
Db 1747 AAAAAAGAACCTGGAATTTATGATGACTTGAAGAGCTTCCAGAGGATCTATTTCT 1806  
QY 1852 GGGGCGCTGAGTGTTCGCTTCAGCGCGCGCGCGCGCGCTGAGATTCGCTACCC 1911  
Db 1807 GGGACTATTGGCTTTTATGCAACAATTCACATATGACTTCAACATGCTATCGGAGC 1866  
QY 1912 ATCTGCTGCGGACGCGCGGAGGCTGCGGCTGCGCGGCGGCTGCTGCTCTCTCC 1971  
Db 1867 AGTGTGTGAAGCAGACCAAGGCAAGCATCGCGCTGCGGCGGCTGCTGCTATCT 1926  
QY 1972 GACCAAGAGAGAGTTCACGAGACCGGTGAAGGCGCGGCGCATGCTACCGCGCTC 2031  
Db 1927 GATCTGAAGAGAGTGTGATGAGTTGTTTAAAGCTAAGGCGGCGCTTGTCCGCTTG 1986  
QY 2032 GACGCG 2037  
Db 1987 CAAGCC 1992

RESULT 12  
US-10-501-282-6651/c  
; Sequence 6651, Application US/10501282  
; Publication No. US20050203280A1  
; GENERAL INFORMATION:  
; APPLICANT: MCWIGHEU, JOHN CALHOUN  
; APPLICANT: ZAGORSKY, ROBERT JOHN  
; APPLICANT: RUSSELL, DAVID PARRISH  
; APPLICANT: FLETCHER, LEAH DIANE  
; TITLE OF INVENTION: ALLOIOCCUS OTTIDIS OPEN READING FRAMES (ORFS) ENCODING  
; TITLE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF



FILE REFERENCE: AM100780 L2  
CURRENT APPLICATION NUMBER: US/10/501,282  
CURRENT FILING DATE: 2004-07-09  
PRIOR APPLICATION NUMBER: 60/333,777  
PRIOR FILING DATE: 2001-11-29  
PRIOR APPLICATION NUMBER: 60/426,742  
PRIOR FILING DATE: 2002-11-18  
PRIOR APPLICATION NUMBER: PCT/US02/36123  
PRIOR FILING DATE: 2002-11-25  
NUMBER OF SEQ ID NOS: 6653  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 6651  
LENGTH: 1754382  
TYPE: DNA  
ORGANISM: *Alloioococcus otitidis*  
US-10-501-282-6651

Query Match 13.0%; Score 267.2; DB 22; Length 1754382;  
Best Local Similarity 48.8%; Pired. No. 1.6e-49;  
Matches 999; Conservative 0; Mismatches 963; Indels 84; Gaps 7;

QY 1 ATGCGACGCTTTCGATCGACAATGACGACTCGTTCACCAACCTGTTCCAGTACATC 60  
DB 7807 ATGAAGTACTCATATGATCGATACTATGATTTCTTATACCTTATACAGCTTATT 7748  
QY 61 GCGAGGCGCAACCGGCAACCCCGTGTGCGCCCAACGAGCGGCACTGTCGCGGCTG 120  
DB 7747 GCGAAGTTCACAGGAAAGAGCCCATGTGTATAAAAATGACAGATGACTTACCAAGAA 7688  
QY 121 CCCGTGCA--GGACTTGAACGGATGTCGTGTCCCGGAGCCCGGACGCCGACCGG 177  
DB 7687 CTATTGATCTCGACTTTGATTAATGTCATTATTTTCAACGGGAGCCGGTATGTCGACCG 7628  
QY 178 GAACGGGACTTCGGAATCAAGCCCGCGGCGATCAACGACAGCGGCTGCGCCGCTCTGCG 237  
DB 7627 GACAAGGACTTCGCTGTGTCCGCGAAGTAATGAAAAGCTTGAACAAGCTTATCTTTGGC 7568  
QY 238 GTGCTCTCGGCGCAACGAGGATCGCCAGCTCTTGGCGGAAACCGTGGCTGCGCCG 297  
DB 7567 ATTTCCTGGGCGCAACGAGGCACTTATTTATTAACGCTGCAACTTATGAGGGGCTGAT 7508  
QY 298 GAACCATGACAGCGCGGCTCTCCGAGTGCAGCAACCGGAGAGAGAGCTTTCCGGGCG 357  
DB 7507 ATCCCATGATGATGTCGCGAGAGTTCGCTTACCATATGGAATAATATCTTTGACGCG 7448  
QY 358 CTCCCTCGGCTTCAACCGCGTGCCTTACCACTCCCTG--CCGCAACGACCTCCCC 414  
DB 7447 ATGAGCCAAAGCTTTGAAGTCAACCGCTACCACTCATTTGTTGTGAAGCAAGAAATTA 7388  
QY 415 GACGAGCTCGAAACCCCTCGCTGGAGGAGCAAGAGGAGTGTGATGAGGCTGCGGCAACCG 474  
DB 7387 GAGAAATTTCAATTCATGTCGCAAAACCGATGATGGAATTTGTCAATGAGCCCTGTCCCAAG 7328  
QY 475 GAGAAGCCGCTGTGGGGGCTGCAAGTTTCACCCGAGTTCATCGGACGCACTTCGCGCG 534  
DB 7327 ACCAAGCCCATATGAGGATCCAAATTCACCCGAAATTCATGTCACCCAGTATGAGGAA 7268  
QY 535 GAATCATGAGCAACTTCGCGGACCTGCGCTGCGCCACCAACGAGGAGAGCGGCGCACGG 594  
DB 7267 AAATATGAGAAAACCTT-----TATGGCTTTGTGCGACAGAC 7232  
QY 595 GCGGACTCCCGTATGCAACTTCACGTCGCGCGTGCAGCGTGCAGCGCGCAAGAG 654  
DB 7231 TACTATTAACCAATTCAGCTTTATTTATGAAAAAGTTCCGGAACTAGGAGCAGCGCAAGAC 7172  
QY 655 GTAGCGCGGAGCTGCTGCGCGGAGGAGCAACGTTTCGTGTCGACAGAGCTCCGTC 714  
DB 7171 CTATAGATCAATTTGGCAAAATGATGACCAAGTACTTTGGCTGACCTCAGAGAGGTG 7112  
QY 715 CTGGAAGCGGCTGCGCTTCTCTCTCGGAGAGCAACGCGGCGCGCTGCGGAGTAC 774  
DB 7111 GAGCGCTGCTTGTACAGGTTTTCATCTTTGCGATGCGTGGACCAAGAGCGGGGCAACCC 7052

QY 775 CTCACCTAACCGGTGCGCGACGCGTGTCTTCGCTCGCGGCTCCGACGCAACGACCC 834  
DB 7051 TTAAAAATAGATGTGACCAAAAGAAAGTAAAAAACAAGATAGCAGTCCGACAGATT 6992  
QY 835 CGAGCGGCGGCGCTTCTTCAACTACCTGAGAGAGCAGCTGGAACGCCGAGGATCCCC 894  
DB 6991 GAAA--GCTTCAAGATGACATCTTTTCTTAAAGGCCAACCGCAAAATGCGCT 6935  
QY 895 GTGCGCCCGGAATGCGCTTTCAGATTCAACTCGGCTACGTCGCTACCTCGGCTACAG 954  
DB 6934 TATGAGCGGCGCTTGTGCTTTCATTTTACGTAAGCTATATTTGGCTATATGCGCTATGAG 6875  
QY 955 CTGAAGCGGAGAC--CACCGGCAACCCCGGCAACCGGTCCCGCAACCCGACGCGCG 1011  
DB 6874 GTCAAAAAAGATATACGTTCAAGTTCACCAACAGACACAGTCACTATCCGATGCTTAC 6815  
QY 1012 TTCTCTTTCGCGGACCGGCGCATGCGCTTCACACACAGAAAGGCTGCTGCTACTGCTG 1071  
DB 6814 TTTACTTATTTGACCGGCGCTTGTGATGACACACAGAAAGAGACCTTTTACTTCTTC 6755  
QY 1072 GCGCTGACCGCGGCGGACAGACGCGCGCGCTGCGTGGGAGACGCGCGAG 1131  
DB 6754 TCTTACCA-----GGATGAC 6740  
QY 1132 ACCCTCACCGGCTGTGCTGCGCGCCCGCGGACCGGACCCCGCATGTGCTTTCGCG 1191  
DB 6739 CAGGATGTGATTCAGCGGTTCAAAAGCAAGGCTCAAGCAGGCGAGTCAAGTTTCAGCAAGA 6680  
QY 1192 ATCCCGAGGCGGCGGCTTTCGCGCTTTCGCGCTTTCGCGCGCGCGCAACAGAGAGCGC 1251  
DB 6679 GGAACAAAGGAGCAAGCAACTTCCACGACTGCGCTTCG--TCAAAACAAACAGCGC 6623  
QY 1252 TACCTCAAGCGGATGACAGAGTGCCTCAAGAGATCCGCAACGCGGAGTGTGACGATTC 1311  
DB 6622 TATATTAAGACATTTTAAACATTCACAGCTGATACAGGCTGCGGAAAGTTATGAAGT 6563  
QY 1312 TGCCTGACCAATGTGTCAACGCGCGGACCGAGGCGGACCGGCTGCTTATCTCCGCG 1371  
DB 6562 TGCCTTACCAACCGGCTGATATTTAGGAGTCAAGTATGAGCTGCTGCTTATTAACAACAC 6503  
QY 1372 CTGCGGCGCATAGCGGCTTCCGTAAGGCGGCTGCTGAGTTTCCGCACTGCTGCGG 1431  
DB 6502 TTGCGGCAAGAAAGTCCCGGTCAGTACTGCTTTTTCCTTTTGTGATATTTTCAAGTT 6443  
QY 1432 CTGAGCGCTCGCGGAGCGGTTTCCACAGTGGGCGGACGCGGCGCTGAGTCAAG 1491  
DB 6442 CTTTCTCTTTCATGAGGCTTTTATTAATGTCAGACAGGACCGGCTTGTGACCACTAAG 6383  
QY 1492 CCCATCAAGGAGACCCGCGCGGCGGCAACCGGAGAGAGAGAGCGGCTCCGCGCG 1551  
DB 6382 CCCATCAAGGAGAGCGGTCAAGAGGAGGAGACAGATGACCAAGAAACCAAGACTTGAAGA 6323  
QY 1552 GACTGCGCGGCGGAGAGAGACCGGCGCGAAGCTGATGATGTGCACTGCTGCGC 1611  
DB 6322 GGCCTCCGAGTGAAGAGAAACCAAGGCGAAGAACTGATATTTGTCACCTCTTGC 6263  
QY 1612 AACGACTCAAGAGGTCGCGGATTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1671  
DB 6262 AATGATCTGAGCGGCTTTTGTGAATCGGCTGCTGAGTACCAAACTGATGATGATGA 6203  
QY 1672 GAGACTTACGCGCGGCTGACAGAGTGTGTGACACATCCGAGGAGCGGCTGCGCGCGCG 1731  
DB 6202 GAGACTTACTCACCCTTCACAGAGCTGTATACAGGTTTCAAGGCGGCTCAAAAGAGAC 6143  
QY 1732 ACCAGACCGCGCTGCTGACAGCGCGCTTTCGCGGCTGCTGCTGCTGCTGCTGCTGCT 1791  
DB 6142 CTATGATGTGTTGAAGTGTAAAGAAACACTTTCGCGGCGGCTGCTGCTGCTGCTGCT 6083  
QY 1792 AAGAAAGCAACATGAGATTCATTCAGCGGCTGAGAGAAAGGCGCGGCGGCTGCTATCC 1851  
DB 6082 AAAAAAGAACCTTGAAATTTATGATGACTTGAAGAGCTTCCAGAGAGATTAATTTCT 6023  
QY 1852 GGGGCGCTCGATGTGTTGCGCTTCAAGCGGCGCGCGGACCTGAGATGTCATCCGACCC 1911



```

Db      6022 GGGACTATTTGGCTTTTATAGCCCAATTCACCTATAGCTTCAACATGTCATCCGAGACC 5963
Qy      1912 ATCTGCTGGCCGACGCGCAGAGTTGGCGTCGGGGGAGTCGTCTCCCTCC 1971
Db      5962 AGTGTGTGGAAGCAGACAAAGCAAGCATCGGCTGGGGGGTGGATTGTCATGCTATCT 5903
Qy      1972 GACCAGAGAGAGAGTTTCAACGAGACCGTGGTAAAGCCCGCATGTATCACCGCCCTC 2031
Db      5902 GATCCTGAAGAAAGATTGATGAGTTGTTTAAAGCTAAGGGGCGCTTGTCCGCTTG 5843
Qy      2032 GACGCG 2037
Db      5842 CAAGCC 5837

```

```

RESULT 13
US-09-738-626-1111
; Sequence 1111, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1111
; LENGTH: 1860
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1111

```

```

Query Match      12.9%; Score 266.8; DB 9; Length 1860;
Best Local Similarity 59.7%; Pred. No. 5.6e-49;
Matches 467; Conservative 0; Mismatches 312; Indels 3; Gaps 1;

Qy      1242 CAAGAGCGCTTACTCAAGCGCATCGACGAGTCTCAAGAGATCCGAAACGGCGAGTC 1301
Db      1059 CAAAGATGAGTATCTCGAATAATTGCGAGAGCCGAGAGCTGATTACTCGCGGCAATC 1118
Qy      1302 GTAACGATCTGCTGACCAACATGCTACCGCGCGGACGAGCGACGGCCCTGCGCT 1361
Db      1119 GTATGAAATCTGCTACCACAAACTTCAAGGCAACACTGATGTCCTCTGCTGCTGC 1178
Qy      1362 CTAAGTGGGCTGCGCGCATCATGAGCCGTCGCTGCGCGGCTGCTGAGTTCCCGCA 1421
Db      1179 CTATCTAGACACTGCGTGGGCGCATCCACCGCATATGCTATCTTCAAGTGGGGA 1238
Qy      1422 ACTGTGGTGTGAGGCGCTCGCGCGAGGGTTCCTCAGATCGCGCGGACGGCGCGCT 1481
Db      1239 TACCTCTATTTTGAAGTTCTCGCGCGAGGGTTCACTCATTTGATTGGCAGGGTATCT 1298
Qy      1482 CGAGTCAAGCCCATCAAGGGACCGCGCGCGCGCGACCGCGGAGAGAGACGAGCG 1541
Db      1299 GGAATCAAGCCCATTAAGGGACACGAGCGCGCTGGGCAACGCGCAAGAGACCAAGA 1358

```

```

Qy      1542 GCTCCGCGCGACCTGCGCGCGGAGAGACCGGCGGAGACCTGATGATCGTGA 1601
Db      1359 AATCATTTGCTGAGCTGCGAGTAACTCTTAAAGATCTGCGAGAAACTTATGATCTGGA 1418
Qy      1602 CTTGTTCGCAAGACCTCAACAGCGTCTGCGGATCGGCTCCGTCACGTCGCCGCT 1661
Db      1419 TTGTGTCGCAAGACCTTAAAGCGCGCGCTTTGGCCACACAGTTTAAACATCAAGCT 1478
Qy      1662 CTTGAGTGAAGACTTACGCGCGCGTGCACACAGCTGTGTGACCATTCGGGAGCGCT 1721
Db      1479 TTTCGACGTGAAACCTTACGCGCAAGTCCACCACTTGTTCAGCACCGTCTTCAGAGTT 1538
Qy      1722 GCGCGCGGACGACGACCGCGCTGCGGTAAGGCGCGCTTCCGCGGCGCTCAATGAC 1781
Db      1539 GGGGCCACG---CAGTTCGATTGATGCTGCGCGAGCATTTCCGAGTGTTCGATGAC 1595
Qy      1782 CGGCGGCGCCCAAGACGACCATGAGATCATCGACCGCTGAGAGAAAGCCCGGAG 1841
Db      1596 TGTGCGCCCAAGCTGCGACCATGAGATCATGATGAGCTGAGGCGAGCTCTCGCG 1655
Qy      1842 CGTCTACTCGGGGCGCTGATGCTTCCGCTCAAGCGCGCGCGACCTCAGATCGT 1901
Db      1656 TATTACTCAGGTGCTGGGATATTTTCCCTCGACGCGCAGTTGATCTCTCATGGT 1715
Qy      1902 CATCGCACCATGCTGCTGCGCGACGCGGCGAGGCTCGGCTGCGGCGCGCATCGT 1961
Db      1716 GATCAGAACTCTGATATCATCAAGACATCATCGTGGATGAGAGTGGGCGGTCACTTCT 1775
Qy      1962 GTCCCTCTCCGACCAAGAGAGAGTTTCAACGAGACCGTGTAAAGGCCCGCGCATGCT 2021
Db      1776 TGCTCTGTGATTCGGAAGCTGAGTGGAGAAATCCCGTTAATACGCGCTCTGCT 1835
Qy      2022 CA 2023
Db      1836 GA 1837

```

```

RESULT 14
US-10-494-675-25
; Sequence 25, Application US/10494675
; Publication No. US20050019877A1
; GENERAL INFORMATION:
; APPLICANT: Zelder, Oskar
; APPLICANT: Pompejus, Markus
; APPLICANT: Schroder, Hartwig
; APPLICANT: Krogger, Burkhard
; APPLICANT: Klopprogge, Corinna
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: Genes coding for metabolic pathway proteins
; FILE REFERENCE: BGI-163US
; CURRENT APPLICATION NUMBER: US/10/494,675
; CURRENT FILING DATE: 2004-05-04
; PRIOR APPLICATION NUMBER: PCT/EP02/12141
; PRIOR FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: DE 101 54 292.1
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 164
; SEQ ID NO 25
; LENGTH: 2005
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1975)
; OTHER INFORMATION: RXA00579
US-10-494-675-25

```

```

Query Match      12.9%; Score 266.8; DB 21; Length 2005;
Best Local Similarity 59.7%; Pred. No. 5.5e-49;
Matches 467; Conservative 0; Mismatches 312; Indels 3; Gaps 1;

Qy      1242 CAAGAGCGCTTACTCAAGCGCATCGACGAGTCTCAAGAGATCCGAAACGGCGAGTC 1301

```



Db 1174 CAAAGATGAGTATCTCGACAAAATTGCGAGAGCCGAGAGCTATTACTCGCGCGAATC 1233  
QY 1302 GTACGAGATCTGCTGCTGACCAATGGTCACTCGGCGGACCGAGCGGACGCTGCGCT 1361  
Db 1234 GTATGAATTTGCTTGAACCAAACTTCAGGGGACACACTGATGTGCCCCCTCTGGTGC 1293  
QY 1362 CTACTCCGCGCTGCGGCGCATCAGCCCGTCCGTACGCGGCGCTGCTGCAATTCGCCGA 1421  
Db 1294 CTATCTAGCACTGCGGGGCGCAATCCACCGCATATGTCGCTATCTTCACTGAGTGGGGGA 1353  
QY 1422 ACTGTGGTCTGAGCGCTTCGCGCCGAGCGGCTTCCTCAGCATGCGCGCGAGCGGCT 1481  
Db 1354 TACCTTATTTTAAAGTTCCTCGCGGAGCGGCTTCATCACCATTGATTCGCGAGGATATG 1413  
QY 1482 CGAGTCCAAAGCCCATCAAGAGGAGACCGCGCCCGGGGCGGACCGCGGAGGAGGACGAGCG 1541  
Db 1414 GGATCAAAAGCCCATTAAGAGCCAGGCGCGCTGGCGCAACGCGGAGAGAGACCAAGA 1473  
QY 1542 GCTCCGCGCGGACCTTGCGCGCGCGGAGAGACCGGCGCGAGAACCTGATGATCGTCA 1601  
Db 1474 AATCATTTGCTGAGCTGGCGAGTAACTCTAAAGATCGGAGAGAAAATTGATGATCGTGA 1533  
QY 1602 CTTGTCCTCGGACGACCTCAACAGCTCTGCGGATGCGCTCCGTCAAGTCCCGGCT 1661  
Db 1534 TTTGTCCTCGGACGACCTTACCGCGCGCTTTGCGGACCAAGTTAAACATTCAGGCT 1593  
QY 1662 CTTGAGGTGAGACCTTACGCGCGCGGCGGACGAGCTGTGACCATCCGGGAGGAGGCT 1721  
Db 1594 TTTGAGCTGAGAACTTACGCGCAAGTCCAACTTTGAGCAACGCTCTCTGCAAGATT 1653  
QY 1722 GCGGCGCGGACGACGACCGCGCTGCTGACGCGCGCGCTTCGCGGCGCTCAATGAC 1781  
Db 1654 GGGGCGACG---CAGTCCGATTGAGTGCCTGCGCGGACGATTCGCGGCTGTGATGAC 1710  
QY 1782 CCGCGCGCGCAAGAGCGCACCTGAGATCATGACCGCTGAGAGAGCGCCCGGCG 1841  
Db 1711 TGTGTGCGGCAAGCTGCGGACCATGAGATCATGAGCTGAGAGCGAGCTCTCGCGG 1770  
QY 1842 CGTCTACTCGGGGCGCTGAGTGTGCGCTCAGGAGGCGCGGACCTCGACATCGT 1901  
Db 1771 TATTTACTCAGGTGCTTGGATATATTTTCTCGAGCGGCGGAGTTGATTCCTCAAGGT 1830  
QY 1902 CATCCGACCATCTGCTGCGCGGACGCGGCGGATTCGCGCTGCGCGGCGATCGT 1961  
Db 1831 GATCAGAACTCTGCTCATCCAGAAATACGTTGAGTACGAGTGGGCGGTGCTCACTTC 1890  
QY 1962 GTCCCTCTCGGACGAGAGAGATTCACCGAGACCGTGTAAAGCGCCCGCATGCT 2021  
Db 1891 TGTCTGTGTGATCCGAGGCTGAGTGGAGAAATCCGCTTAAATCAAGGCTCTGCT 1950  
QY 2022 CA 2023  
Db 1951 GA 1952

RESULT 15  
US-09-738-626-1  
; Sequence 1, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626

; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 1  
; LENGTH: 3309400  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-1

Query Match 12.9%; Score 266.8; DB 9; Length 3309400;  
Best Local Similarity 59.7%; Pred. No. 1.8e-49;  
Matches 467; Conservative 0; Mismatches 312; Indels 3; Gaps 1;

QY 1242 CAAAGAGGCTACTCTCAAGCGCATCGAGTGGCTCAAGAGATCCGCAAGCGGAGTC 1301  
Db 1053079 CAAAGATGAGTATCTCGACAAAATTGCGAGAGCCGAGAGCTGATTACTCGGCGAATC 1053138  
QY 1302 GTACGAGATCTGCTGACCAACATGCTCAACGCGCGGACCGAGCGGACCGCTGCGCT 1361  
Db 1053139 GTATGAATCTGCTGACCAAAAATTCAAGGACCACTGATGTGCCCCCTCTGGCTGC 1053198  
QY 1362 CTACTCCGCGCTGCGGCGCATCAGCCCGCTCCGTACGCGCGCTGCTGATTCGCCGA 1421  
Db 1053199 CTATCTAGACCTGCGGGGCGCAATCCACCGCATATGCTGTGATCTTCACTGGGGA 1053258  
QY 1422 ACTGTGGTCTGAGGCGCTCGCGCGGAGGTTCTCAGATCGGCGCGGAGCGGCGCT 1481  
Db 1053259 TACCTTATTTTGAAGTCTCTCGCGGAGCGGTTCACTGATGATTCGCGAGGATAT 1053318  
QY 1482 CGAGTCCAAAGCCCATCAAGAGGAGACCGCGCCCGGCGGACCGCGGAGAGAGACGAGCG 1541  
Db 1053319 GGATCAAAAGCCCATTAAGAGCACGAGCGCGTGGCGCAACGCGCAAGAGACCAAGA 1053378  
QY 1542 GCTCGCGCGGACCTGCGCGCGCGGAGAGAGAGCGGCGGAGAACTGATGATCGTCA 1601  
Db 1053379 AATCATTTGCTGAGCTGCGGAGTATCTTAAATGCTGAGAGAACTTGATATCGTGA 1053438  
QY 1602 CTTGTCCTCGGACGACCTCAACAGCGCTCGCGGATCGGCTCGTCCAGTCCCGGCT 1661  
Db 1053439 TTTGTCCTCGGACGACTTACCGCGCGGCTTTGCCACCAAGTTAAACATTCAGGCT 1053498  
QY 1662 CTTGAGGTGAGACCTTACGCGCGCGTGCACCAAGCTGTGTGACCATCCGGGAGCGCT 1721  
Db 1053499 TTTGAGCTGAGAACTTACGCGGACAGTCCACCACTTGTCAAGACGCTCTGAGAGATT 1053558  
QY 1722 GCGGCGCGGACCAAGACCGCGCTGCTGAGCGCGCGCTTCGCGGCGCTCTCATGAC 1781  
Db 1053559 GGGGCGACG---CAGTCCGATTGAGTGCCTGCGGAGCATTCGCCGAGTTCGATGAC 1053615  
QY 1782 CCGCGCGCGCAAGAGCGGACCATGAGATCATCGACCGCTGAGAGAGGCGCCCGGCG 1841  
Db 1053616 TGTGCCCCCAAGCTGCGGACATGAGATCATCGATGAGCTGAGAGGACGCTCTGCGG 1053675  
QY 1842 CGTCTACTCGGGGCGCTGAGTGTGCGCTTCAGCGCGCGCGGACCTGAGATCGT 1901  
Db 1053676 TATTTACTCAGGTGCTTGGATATTTTCTCGAGCGGCGGAGTTGATCTCTCATGCT 1053735  
QY 1902 CATCCGACCATGCTGCTGCGGACGCGGCGGAGTTGCGCTGCGCGGCGATCGT 1961  
Db 1053736 GATCAGAACTCTGCTCATCCAGAAATACATCAGTGGAGTACGAGTGGGCGGTGCACTTCT 1053795  
QY 1962 GTCCCTCTCGGACGAGAGAGAGTTCACGAGACCGTGTAAAGCGCGCGCATGCT 2021  
Db 1053796 TGTCTGTGTGATCCGAGGCTGAGTGGAGAGAAATCCGCTTAAATCAAGGCTCTGCT 1053855  
QY 2022 CA 2023



Thu Oct 6 09:19:05 2005

us-10-089-514-1.rnpb

Page 17

Db 1053856 GA 1053857

Search completed: October 5, 2005, 09:56:39  
Job time : 2072.38 secs

---



**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 20:20:08 ; Search time 7793.81 Seconds  
(without alignments)  
1065.737 Million cell updates/sec

Title: US-10-089-514-1

Perfect score: 2061  
Sequence: 1 atgcgcacgcttcgcatcga.....ccgtggcggcgcccgatga 2061

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68473088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:\*  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	216.6	10.5	529	CA003778	CA003778 HS15114r
2	212.8	10.3	845	BZ570522	BZ570522 msh2_1411
3	205.8	10.0	653	CF863831	CF863831 p82S008xg
4	195.6	9.5	1043	BZ548859	BZ548859 pasc1-60
5	178	8.6	523	BZ412600	BZ412600 OGAAG11TM
6	178	8.6	631	BZ412593	BZ412593 OGAAG11TC
7	178	8.6	631	BZ537360	BZ537360 OGAAG11TC
8	178	8.6	810	BZ537360	BZ537360 OGAAG11TC
9	174.2	8.5	579	CF598037	CF598037 NCST3e06
10	171.8	8.3	660	BZ892725	BZ892725 Hm7_0152
11	167.8	8.1	894	CO024641	CO024641 EST803025
12	167.2	8.1	757	BZ896109	BZ896109 NARP6_013
13	164.8	8.0	750	CO028609	CO028609 EST806993
14	164.8	8.0	820	CO030342	CO030342 EST808726
15	164.8	8.0	820	CO034510	CO034510 EST812894
16	163	7.9	768	CO980064	CO980064 GM89007A1
17	162.8	7.9	859	CC144065	CC144065 NDL_75A11
18	162.8	7.9	859	BZ565136	BZ565136 GM700005A
19	160.6	7.8	1674	CL078342	CL078342 CH216-149
20	159.6	7.7	2041	AG365808	AG365808 Mus muscu
21	158.6	7.7	1259	BZ557654	BZ557654 pasc1-60
22	158.4	7.3	432	CV122440	CV122440 Md1v4003E
23	151.2	7.3	1695	CC290874	CC290874 CH261-172
24	150.6	7.3	1031	BZ561643	BZ561643 pasc2-164

25	150.4	7.3	492	6	CB518127	CB518127 EST01 DNA
26	148.2	7.2	528	2	AW509018	AW509018 s139b01.Y
27	148.2	7.2	698	5	B0625203	B0625203 USDA-PP 0
28	147	7.1	1628	9	CG757066	CG757066 P052-2-10
29	145.4	7.1	1450	9	AG429738	AG429738 Mus muscu
30	143.8	7.0	1332	9	AG363333	AG363333 Mus muscu
31	141	6.8	1821	9	CL090560	CL090560 ISB1-17N1
32	139.6	6.8	1523	9	AG448267	AG448267 Mus muscu
33	139.2	6.8	1567	6	CD476709	CD476709 P040-1-B0
34	139	6.7	716	6	CD407015	CD407015 Gm_CK3199
35	138.8	6.7	1956	9	CG754548	CG754548 P050-1-A1
36	138.4	6.7	565	6	CA640597	CA640597 wren1.p00
37	138.4	6.7	669	4	B1955833	B1955833 HVSmem02
38	138.2	6.7	1822	9	AG435170	AG435170 Mus muscu
39	138.2	6.7	2243	9	AG381986	AG381986 Mus muscu
40	137.6	6.7	526	7	CF887023	CF887023 UI-CF-D01
41	136.2	6.6	540	4	B1427109	B1427109 sah76601
42	136.2	6.6	1793	9	CG754612	CG754612 P050-1-D0
43	136	6.6	1738	9	CG750956	CG750956 P045-2-H0
44	136	6.6	1830	9	CL078620	CL078620 CH216-151
45	135	6.6	1642	9	CG754720	CG754720 P050-2-A0

## ALIGNMENTS

RESULT 1  
CA003778  
LOCUS HS15114r 529 bp mRNA linear EST 23-OCT-2002  
DEFINITION HS15114r HS Hordeum vulgare subsp. vulgare cDNA clone HS15114  
ACCESSION CA003778  
VERSION CA003778.1 GI:24280760  
KEYWORDS EST.  
SOURCE Hordeum vulgare subsp. vulgare  
ORGANISM Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Hordeum.  
1 (bases 1 to 529)  
Zhang,H., Potokina,E., Michalek,W., Weschke,W., Stein,N. and Graner,A.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Barley ESTs from germinating seeds  
Unpublished (2002)  
Contact: Stein Nils  
Molecular Markers Group, Department Genbank  
Institute of Plant Genetics and Crop Plant Research (IPK)  
Corrensstr. 3, 06466, Gatersleben, Germany  
Tel: 039482-5522  
Fax: 039482-5595  
Email: stein@ipk-gatersleben.de  
Insert length: 529 Std Error: 0.00  
Plate: 15 row: 1 column: 14  
Seq primer: M1rev  
Location/Qualifiers  
1. 529  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="barke"  
/sub\_species="vulgare"  
/db\_xref="GABI:258990"  
/db\_xref="taxon:112509"  
/clone="HS15114"  
/issue\_type="embryo + scutellum"  
/dev stage="0-16 hours after imbibition"  
/lab\_host="X110-Gold"  
/clone\_lib="HS"  
/note="Vector: pBluescript SK+, Site\_1: EcoRI (5'-end of cDNA); Site\_2: XhoI (3'-end of cDNA); Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also



## ORIGIN

due to the cloning system used Blue/white selection for recombinants is not 100% reliable."

Query Match 10.5%; Score 216.6; DB 6; Length 529;

Best Local Similarity 63.1%; Pred. No. 3.7e-27; Mismatches 33; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

```
QY 1482 CGAGTCCAGCCCATCAAGGGAGACCCGCCCCGGGGGGGACACCGGAGAGAGACGAGC 1541
    |||
DB 1 CGGACAGAGGCCCAATCAAGATACATAGACAGTGGGAGAAACACAGAGAGAGATGAGT 60
QY 1542 GCTCCGCGCCGACCTCGCGCGCGGAGAGAGACCGGCGGAGAACTGATGATCGTGA 1601
    |||
DB 61 TCTACGTTTCAACTGAAATATACAGTGAAGAGACAGGCTGAGAACTTGATGATTTGTA 120
QY 1602 CTTGTCGCGACGAGCTCAACAGCGTTCGCGGATGAGCTCGGTCCAGTGGCCCCGCT 1661
    |||
DB 121 TCTCTTAAGAAATGATGTGGGTAAAGTCTGAGCCTGGAGAGCGTGCATGTTCTCGCT 180
QY 1662 CTTGAGAGTGAAGACCTACGCGCCGCGGACCAAGCTGTGTCAGCCATCCGGGAGCGCT 1721
    |||
DB 181 CATGAGAGTGAATCATATATATCTGTTCAACCATGATGATGACACCATCCGCGAACA 240
QY 1722 GCGGCGCGGACGACGACCGCGCGCTGATACGCGCGCTTCCCGCGCGCTCATGAC 1781
    |||
DB 241 GAAGCTGATCTAAGCCCGGTAGACTGATCAAAAGCCGCTTTCAGAGAGGCTCGATGAC 300
QY 1782 CCGCGCGCCCAAGAGGACCATGAGATCATGACCGCGCTGAGAGAGCCCGCGGG 1841
    |||
DB 301 GGGTGCCCCCAAGGTCAAGTGAATGAGATCTCGAGCGCTTGAAGAGCAAGCCCAAGGG 360
QY 1842 CGTCTACTCCGGGGGCGCTCGGATGTTGCGCCCTCAGCGCGCGCGCGACTCGACATCGT 1901
    |||
DB 361 AATATACTCGGGGTCATCGGGTTCTTCTGTACACCGCATCGTTGATTCGAATCGT 420
QY 1902 CATCCGACCATCGTGTGCGCGACGCGCGAGATTGCGCGCTCGCGCGCGCGATCGT 1961
    |||
DB 421 GATCAGAGACGATGAGTGTGACGCGGAGTGGCGCTCGGTGAGAGAGCGGGGCGATTTG 480
QY 1962 GTCCCTTCCGACGAGAGAGATTACCGAGACCGTGTGTAAGGC 2009
    |||
DB 481 GGGCTGTGACACCGGAGGCGAGTACGCCGAGATGATCTCAAGGC 528
```

RESULT 2 B2570522 845 bp DNA linear GSS 17-DEC-2002

LOCUS msh2.1411.xl msh Pseudomonas aeruginosa genomic clone msh2\_1411, genomic survey sequence.

ACCESSION B2570522  
VERSION B2570522.1 GI:27205583

KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa

REFERENCE 1 (bases 1 to 845) Bacteriophage Pseudomonas aeruginosa; Pseudomonadales; Pseudomonadaceae; Pseudomonas.

AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.

TITLE Whole-Genome-Sequence Variation among multiple isolates of Pseudomonas aeruginosa library

JOURNAL J. Bacteriol. (2002) In press

COMMENT Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu

FEATURES  
Source 1..845 Location/Qualifiers

/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="MSH"  
/db\_xref="taxon:287"  
/clone="msh2\_1411"  
/clone\_11b="msh"  
/note="Environmental isolate. Whole genomic shotgun library."

Query Match 10.3%; Score 212.8; DB 8; Length 845;

Best Local Similarity 60.6%; Pred. No. 1.6e-26; Mismatches 349; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

```
QY 1486 TCCAAAGCCCATCAAGGGAGACCCGCCCCGGGGGGGACACCGGAGAGAGACGAGCGCTC 1545
    |||
DB 69 TCGAATTCCTGCAAGCCCAACCGCCCGCGGCAAGACCCCGAGAGAGACATGGGCTTG 128
QY 1546 CCGCGCAGACTGCGCGCGCGGAGAGAGACCGGCGCGAGAACCTGATGATCGTCACTTG 1605
    |||
DB 129 GCGGCGCTGCTGTGCGCGAGCCCGCAAGACCGGCGGAGAAACCTGATGATCGTCACTTG 188
QY 1606 GTCCGCAAGACCTCAACAGCGCTGCGGATGCTCGTCCGACGCTGCGCGCGCTTTC 1665
    |||
DB 189 CTGCGCAAGACATCGAAGCGAGTTCGCAACTGCGGAGGATCGGCTCAAGAGCTGTTG 248
QY 1666 GAGGTGAGACCTACGCGCGCGCGGACGACGAGCTGTGTGACCATCCGGGAGCGGCTGGG 1725
    |||
DB 249 GCCCTGAAAGCTATTCGAAGCTGATCATCTGTGTGAGAGAGCTCACCGGCACTGGCG 308
QY 1726 CCGGAGCAGAGACCGCGCGCTGCGTACGCGCGCGCTTCCCGCGCGCTCATGACCGCG 1785
    |||
DB 309 CCGGAGCAAGAGCGCTCGACCTGACCTGCGAGAGAGAGCTTCCCGGCGGCTCATACCGCG 368
QY 1786 GCGCCCAAGAGAGGACCATGAGATCATGACCGCTTGAAGAGAGAGAGAGAGAGAGAG 1845
    |||
DB 369 GCGCCCAAGATTCGCGCGCATGAGATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 428
QY 1846 TACTCGGGGGCGGTGGAATGATTCGCGCGTCAAGGAGGCGCGGACCTCGACATCGTCATC 1905
    |||
DB 429 TACTGCGGAGCGCTGTTTCTTACTCTGACGTGCGGGGAGAGATGACATCGATCGCATC 488
QY 1906 CGCACCATCGTGTGCGCGAGCGCGGAGATTGCGCGCTGCGCGCGCGATCGTGTCC 1965
    |||
DB 489 CGCACCCTGCTGTGTCAGAAACGCGCAGGTCAATGTGTGGGGGCGGCGGATCGTGGCG 548
QY 1966 CTCTCCGACCAAGAGAGAGATTACCGAGACCGTGTGTAAGGCCCGCCCATGTGTACC 2025
    |||
DB 549 GACTCGACCTGGGAGAGACGATCAAGAAACCTTGAGACGAGGTGCGGTGTGTGGAA 608
QY 2026 GCCCTGAGAGGAGCGCGCGTGGGCGGCGCGCGATGA 2061
    |||
DB 609 ACCCTGAAAGGAGAGGCGCGGACGAGGTTTCCGCGGA 644
```

RESULT 3 CF863831 653 bp mRNA linear EST 31-OCT-2003

LOCUS ps25008xg04f USDA-IFAPs:Expression of Phytophthora sojae genes during infection and propagation\_s25 Phytophthora sojae cDNA clone

ACCESSION CF863831  
VERSION CF863831.1 GI:38118457

KEYWORDS EST.  
SOURCE Phytophthora sojae  
ORGANISM Phytophthora sojae

REFERENCE 1 (bases 1 to 653) Eukaryota; Stramenopiles; Oomycetes; Pythiales; Pythiaceae; Phytophthora.

AUTHORS Tyler,B. Not Published

TITLE Tyler,B. Not Published

JOURNAL Unpublished (2003)  
COMMENT Contact: Tyler B



## Tyler lab

VBI  
1880 Pratt Dr., Blackeburg, VA 24061, USA  
Tel: 540-231-7318  
Email: bmttyler@vt.edu

## PCR Primers

FORWARD: BK reverse primer  
BACKWARD: BK reverse primer  
Plate: 008 row: G column: 04  
Seq primer: BK reverse primer  
High quality sequence stop: 653.

## FEATURES

## source

1. 653  
/organism="Phytophthora sojae"  
/mol\_type="mRNA"  
/db\_xref="taxon:67593"  
/clone="szs008G04"  
/issue\_type="zoospores"  
/cell\_line="P6497"  
/dev\_stage="Free swimming"  
/clone\_lib="USDA-IPAFS: Expression of Phytophthora sojae genes during infection and propagation, szs"  
/note="Vector: PBK-CMV; Site\_1: EcoRI; Site\_2: XhoI"

## ORIGIN

Query Match 10.0%; Score 205.8; DB 7; Length 653;  
Best Local Similarity 58.3%; Pred. No. 2.6e-25;

Matches 379; Conservative 0; Mismatches 268; Indels 3; Gaps 1;

1280 AGGAGATCCGCAACGGGAGTCTGACGAGATCTGCTGACCAATGTCACCGCGCGA 1339  
1280 AGGAGATCCGCAACGGGAGTCTGACGAGATCTGCTGACCAATGTCACCGCGCGA 1339  
4 ACAGAGCGCGGACATCTTCCAGGCGGTGCGCTCCAGCGGCTGCTGAGACCTCCGA 63  
4 ACAGAGCGCGGACATCTTCCAGGCGGTGCGCTCCAGCGGCTGCTGAGACCTCCGA 63  
1340 CCAGAGCGAGCGCCCTGCTCTACTCCGCGCTGCGCGCCATCAGCCCGTCCGTACG 1399  
1340 CCAGAGCGAGCGCCCTGCTCTACTCCGCGCTGCGCGCCATCAGCCCGTCCGTACG 1399  
64 AGAGCGCACCCCGCTGATCTGTACCGCAGATGCGCGCATCAACCCGTCGCCGTACA 123  
64 AGAGCGCACCCCGCTGATCTGTACCGCAGATGCGCGCATCAACCCGTCGCCGTACA 123  
1400 GCGCCCTGCTCGA---GTCCCGCACTGTCCGCTGAGAGCGCTGCGCGCGCGGTC 1456  
1400 GCGCCCTGCTCGA---GTCCCGCACTGTCCGCTGAGAGCGCTGCGCGCGCGGTC 1456  
124 TGTCTTCTCTGACATGAGGAGAGACTTCCAGATCGTGCGGCTTCCCGAGATGCTCG 183  
124 TGTCTTCTCTGACATGAGGAGAGACTTCCAGATCGTGCGGCTTCCCGAGATGCTCG 183  
1457 TCACGATCGGCGCGGCGGCGGTGAGTCCAGGCCCATCAAGGGAGACCGCGCGCG 1516  
1457 TCACGATCGGCGCGGCGGCGGTGAGTCCAGGCCCATCAAGGGAGACCGCGCGCG 1516  
184 TCAAGGTGACATGACCGGTGTGTGAGAGACGACCCGATTTGCCGTGACACGCCACCTG 243  
184 TCAAGGTGACATGACCGGTGTGTGAGAGACGACCCGATTTGCCGTGACACGCCACCTG 243  
1517 GCGGCAACCGGAGAGAGAGAGAGAGCGGCTCCGCGCGCACTGCGCGCGCGAGAGAGAC 1576  
1517 GCGGCAACCGGAGAGAGAGAGAGAGCGGCTCCGCGCGCACTGCGCGCGCGAGAGAC 1576  
244 GCGCCATGACGAGAGAGAGAGAGAGCACTGTGAAGGACCTGCTGCGGAGAGAGAGAG 303  
244 GCGCCATGACGAGAGAGAGAGAGAGCACTGTGAAGGACCTGCTGCGGAGAGAGAG 303  
1577 GGGCCGAGAACCTGATGATGTCGACCTGTGTCGCAACGACCTCAACAGGCTCTGCGGCA 1636  
1577 GGGCCGAGAACCTGATGATGTCGACCTGTGTCGCAACGACCTCAACAGGCTCTGCGGCA 1636  
304 GCGCCGAGAACCTGATGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 363  
304 GCGCCGAGAACCTGATGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 363  
1637 TCAGCTCCGTCGACATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1696  
1637 TCAGCTCCGTCGACATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1696  
364 CCGGATCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 423  
364 CCGGATCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 423  
1697 TGGTGTGACCATTCGCGGAGCGGCTGCGCGCGAGACGACGACCGCGCTGTGTCGCG 1756  
1697 TGGTGTGACCATTCGCGGAGCGGCTGCGCGCGAGACGACGACCGCGCTGTGTCGCG 1756  
424 TCGTGTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 483  
424 TCGTGTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 483  
1757 CCGCTTCCCGCGGCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1816  
1757 CCGCTTCCCGCGGCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1816  
484 CATGTTCCCTGCGGAGCGCTGTCCGCGCGCCCTTAAGGTGTCGTCGTCGTCGTCGTCG 543  
484 CATGTTCCCTGCGGAGCGCTGTCCGCGCGCCCTTAAGGTGTCGTCGTCGTCGTCGTCG 543  
1817 ACCGCTGAGAGAGAGCGCGCGCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1876  
1817 ACCGCTGAGAGAGAGCGCGCGCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1876  
544 GCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 603  
544 GCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 603  
1877 GCGGCGCGCGGACCTGACGATGTCATTCGACACCATGTCGTCGTCGTCGTCGTCGTCG 1926  
1877 GCGGCGCGCGGACCTGACGATGTCATTCGACACCATGTCGTCGTCGTCGTCGTCGTCG 1926  
604 CCGGCTTCCCTGCGGAGCGCTGTCCGCGCGCCCTTAAGGTGTCGTCGTCGTCGTCGTCG 653  
604 CCGGCTTCCCTGCGGAGCGCTGTCCGCGCGCCCTTAAGGTGTCGTCGTCGTCGTCGTCG 653

## RESULT 4

## B2548859

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

## Location/Qualifiers

## 1. 1043

## /organism="Pseudomonas aeruginosa"

## /mol\_type="genomic DNA"

## /strain="1-60"

## /db\_xref="taxon:287"

## /clone="pac1-60 1509"

## /clone\_lib="pac1-60"

## /note="clinical isolate 1-60 Whole genomic shotgun library."

## ORIGIN

Query Match 9.5%; Score 195.6; DB 8; Length 1043;  
Best Local Similarity 62.4%; Pred. No. 1.4e-23;

Matches 306; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

1519 GGGACCGCGGAGAGAGAGAGAGAGCGGCTCCGCGCGACCTGCGCGCGAGAGAGAGCGG 1578  
1519 GGGACCGCGGAGAGAGAGAGAGAGCGGCTCCGCGCGACCTGCGCGCGAGAGAGCGG 1578  
61 GCGACCTCCGAGAGAGAGAGAGAGCGGCTGCGCGCGCTGCTGCGCGAGAGAGAGCGG 120  
61 GCGACCTCCGAGAGAGAGAGAGAGCGGCTGCGCGCGCTGCTGCGCGAGAGAGAGCGG 120  
1579 GCGGAGAACCTGATGATGTCGACCTGTGTCGCAACGACCTCAACAGGCTCTGCGGATC 1638  
1579 GCGGAGAACCTGATGATGTCGACCTGTGTCGCAACGACCTCAACAGGCTCTGCGGATC 1638  
121 GCGGAGAACCTGATGATGTCGACCTGTGTCGCAACGACCTCAACAGGCTCTGCGGATC 180  
121 GCGGAGAACCTGATGATGTCGACCTGTGTCGCAACGACCTCAACAGGCTCTGCGGATC 180  
1639 GGGCTCGTCGACATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1698  
1639 GGGCTCGTCGACATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1698  
181 GCGAGGTGACGAGTACCGGAGCGGCTGCGCGCGAGAGAGTATCCCAAGCTGATACCTG 240  
181 GCGAGGTGACGAGTACCGGAGCGGCTGCGCGCGAGAGAGTATCCCAAGCTGATACCTG 240  
1699 GTGTGACCATTCGCGGAGCGGCTGCGCGCGAGACGACGACCGCGCTGTGTCGCGCG 1758  
1699 GTGTGACCATTCGCGGAGCGGCTGCGCGCGAGACGACGACCGCGCTGTGTCGCGCG 1758  
241 GTGAGAGAGTACCGGAGCGGCTGCGCGCGAGAGAGCGGCTGTCGTCGAGAGAGCG 300  
241 GTGAGAGAGTACCGGAGCGGCTGCGCGCGAGAGAGCGGCTGTCGTCGAGAGAGCG 300  
1759 GCGCTTCCCGCGGCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1818  
1759 GCGCTTCCCGCGGCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1818  
301 AGCTTCCCTGCGGAGCGCTGTCCGCGCGCCCTTAAGGTGTCGTCGTCGTCGTCGTCG 360  
301 AGCTTCCCTGCGGAGCGCTGTCCGCGCGCCCTTAAGGTGTCGTCGTCGTCGTCGTCG 360  
1819 CCGCTGAGAGAGAGCGCGCGCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1878  
1819 CCGCTGAGAGAGAGCGCGCGCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1878  
361 GAGCTGAGAACCGAGCGGAGCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 420  
361 GAGCTGAGAACCGAGCGGAGCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 420  
1879 GCGGCGCGCGGACCTGACGATGTCATTCGACACCATGTCGTCGTCGTCGTCGTCGTCG 1938  
1879 GCGGCGCGCGGACCTGACGATGTCATTCGACACCATGTCGTCGTCGTCGTCGTCGTCG 1938  
421 GCGGAGTGTGACGATGTCGATTCGACACCATGTCGTCGTCGTCGTCGTCGTCGTCGTCG 480  
421 GCGGAGTGTGACGATGTCGATTCGACACCATGTCGTCGTCGTCGTCGTCGTCGTCGTCG 480  
1939 TTCCGCGTCCGCGGAGCGGCTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1998  
1939 TTCCGCGTCCGCGGAGCGGCTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1998



Db	481	TGCTGGGGGGGGGGGGGATGTCGCCGACTTCGACCTGGAGAGACGATCAGGAAACC	540
Qy	1999	GTGTTAAGG	2008
Db	541	CTGGACAAAG	550
RESULT 5			
B2412600/c			
LOCUS	B2412600	523 bp	DNA
DEFINITION	OGAAG11TM ZM 0.7_1.5_KB zea mays genomic clone ZMMBMA0015004,	linear	GSS 04-DEC-2002
ACCESSION	B2412600		
VERSION	B2412600.1	GI:26047761	
KEYWORDS	GSS.		
SOURCE			
ORGANISM	Zea mays		
	Zea mays		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD		
	clade; Panicoideae; Andropogoneae; Zea.		
REFERENCE	1 (bases 1 to 523)		
AUTHORS	WhiteJaw,C.A., Quackendush,J., Van Aken,S., Uteck,T.,		
	Reenick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,		
	Cieck,R.W., Nurnberg,A., Robbins,D. and Lakey,N.		
TITLE	Consortium for Maize Genomics		
JOURNAL	Unpublished (2002)		
COMMENT	Other GSSs: OGAAG11NC		
	Contact: Cathy WhiteJaw		
	TIGR		
	9712 Medical Center Drive, Rockville, MD 20850, USA		
	Tel: 301-838-5843		
	Fax: 301-838-0208		
	Email: whiteJaw@tigr.org		
	Seq primer: 7R		
	Class: sheared ends.		
FEATURES			
source	location/Qualifiers		
	1..523		
	/organism="Zea mays"		
	/mol_type="genomic DNA"		
	/strain="B73"		
	/db_xref="taxon:4577"		
	/clone="ZMMBMA0015004"		
	/clone_11b="ZM_0.7_1.5_KB"		
	/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb		
	methylation filtered genomic DNA library"		
ORIGIN			
Query Match	8.6%;	Score 178;	DB 8; Length 523;
Best Local Similarity	61.8%;	Fred. No. 1.4e-20;	
Matches 283; Conservative	0;	Mismatches 175; Indels	0; Gaps 0;
Qy	1559	CCGGCCGGGAGAAAGGACCGGGCCGAGAACCTGATGATCGTGCACCTGGTCCGCAAGACC	1618
Db	509	CCACCAGGAAAAAGACCCAGGCCGAGAACCTGATGATGTCGACTCTTAAAGAAACGATC	450
Qy	1619	TCAACAGCGTCTGCGCGATCGGCTTCGTCACGTCGCCCGGCTCTTCGAGTGAAGACT	1678
Db	449	TGGCAAGGTCTGCGAGCCAGGAGACCGTCATGTCCTCACGCTCATGAGCGTTGATCAT	390
Qy	1679	ACGGCCCGGTGACCAAGCTGTGTGTGACCAATTCGGGGACGGCTGGCGCCCGGACACGCA	1738
Db	389	ACAGAAAGTTTCACACCAATGGTGAATACCTCTGGGAAACAAAGACCAATCTMAACC	330
Qy	1729	CCGCGCGCTGGGTAGCGCGCGCTTCCCGGGGGCTCCATGACCGGGCGGCGCAAGAC	1798
Db	329	CCGTTGATTTGTGTAAAGCGGCTTTCCAGGGCGTTGCATGACCGGGGCTCCAAAAGTCA	270
Qy	1799	GCACCATGAGATCATTCGACCGCTCGAGGAAGGCCCGCGGGCGCTCTACTTCGGGGCGC	1858
Db	269	GATCATATGGAATCTTATTTACATTGAGACAAGCCCAAAGAGATGTACTCAAGATGA	210
Qy	1859	TCGATGTTGCGCCTTCACGGCGCGCGGACCTCAGCATGTCATCCGACCATCGTGC	1918

Db	209	TCGGGTTCTTCTCTGTACACCAACCACTTCGACCTCGAACATTGTGATTCGAACTGTATNC	150
Qy	1919	TGCGCGACGGCCGACGGAAGTTTCGGCGTGGCGGGCGATCGTGTCCCTTCGACGAG	1978
Db	149	TGCACGACGGAAGAGCTTCGGTGGGGCGGGTGGGGCGATCGTACGCGCTGTGACCCAG	90
Qy	1979	AGGAGGAGTTTACCGGACCGTGGTAAAGCCCGCCGC	2016
Db	89	AAGCAGAGTACGACGAAATGCTGCTGAAGCAAGCC	52
RESULT 6			
BZ412593			
LOCUS	BZ412593	631 bp	DNA
DEFINITION	OGAAG11TC ZM_0.7_1.5_KB Zea mays genomic clone ZMMBma0015004,	linear	GSS 04-DEC-2002
ACCESSION	BZ412593		
VERSION	BZ412593.1	GI:26047739	
KEYWORDS	GSS.		
SOURCE	Zea mays		
ORGANISM	Zea mays		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoidae; Andropogoneae; Zea.		
REFERENCE	1 (bases 1 to 631)		
AUTHORS	Whitefaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Numbers,A., Robbins,D. and Lakey,N.		
	Consortium for Maize Genomics		
	Unpublished (2002)		
TITLE	Other_GSSs: OGAAG11TM		
JOURNAL	Contact: Cathy Whitefaw		
COMMENT	TIGR		
	9712 Medical Center Drive, Rockville, MD 20850, USA		
	Tel: 301-838-5843		
	Fax: 301-838-0208		
	Email: whitefaw@tigr.org		
	Seq primer: TP		
	Class: sheared ends.		
FEATURES	Location/Qualifiers		
source	1..631		
	/organism="Zea mays"		
	/mol_type="genomic DNA"		
	/strain="B73"		
	/db_xref="taxon:4577"		
	/clone="ZMMBma0015004"		
	/clone_lib="ZM_0.7_1.5_KB"		
	/note="vector: pBESK-; Site_1: HincII; 0.7-1.5 kb		
	methylation filtered genomic DNA library"		
ORIGIN			
Query Match	8.6%; Score 178; DB 8; Length 631;		
Best Local Similarity	8.8%; Pred. No. 1,4e-20;		
Matches 283; Conservative	0; Mismatches 175; Indels	0; Gaps	0;
Qy	1559	CCGGCCGGGAGAGACCGGGCCGAGAACCTGATGATGTCGACCTGGTCCGACGAC	1618
Db	15	CCACCAAGGAAAAAGACCGAGCCGAGAACCTGATGATGTCGACCTCTTAAGAAACGATC	74
Qy	1619	TCAACAGCGTTGGCGCATGGCTCCGTCCACGTGCCCCGGCTCTTGAAGTGGAGACT	1678
Db	75	TCGGCAAGGTCTGGAGCCAGGAGCGGTGATGCCACGCTTATGAGACGTTGAATCAT	134
Qy	1679	ACGGCGCCGTCACACAGCTGGTGTGCAACATCCGGGAGCGGCTCGGGCCCGACACAGA	1738
Db	135	ACAAAGTGTTCACACCATGTGTGATACCGTTCTCGGAACAAGAACGACATTTAAGCC	194
Qy	1739	CCGCGCGCTCGGTACGGCGCGCTTCCCGGGCGGCTCCATGACCGGCGCGCCCAAGAGC	1798
Db	195	CCGTTGATTTGTGTAAGCGGCGCTTTTCCAGGCGGTTGATGACCGGGGCTCCAAAAGTCA	254
Qy	1799	GCACCATGAGATCATGACCGCTTGAGAGAGGCCCCGGGGCGTCTACTCCGGGCGCC	1858



Db 255 GATCAATGAGATCTTGAATCACTTGAGACAAAGCCAGAGAGATGTAATCAGATCGA 314

Qy 1859 TCGGATGTTTCGCGCTCAGCGCGCGCCGACCTCAGATGTCATCCGACCATCTGTC 1918

Db 315 TCGGATGTTTCGCTGTAACAACACATTCGACCTGAAATTGTGATCGAATCTGTATCC 374

Qy 1919 TGGCCGACGCGCAGCGAGAGTTTCGCGCTCGCGCGCGAGATGTCCTCTCCGACGAG 1978

Db 375 TGCACGACGAGGAAGCTCGGTCGCGCGCGAGTGTGAGCGCTGTGAGCGCCAG 434

Qy 1979 AGGAGAGTTTACCGAGACCGTGTAAAGCCCGCC 2016

Db 435 AAGCAGAGTACGACGAATGCTCTGAAAGCGAAAGCC 472

RESULT 7  
BZ537360 631 bp DNA linear GSS 16-DEC-2002  
LOCUS OCAGD95TC\_ZM2\_0.7\_1.5\_KB\_Zea\_mays\_genomic\_clone\_ZM58Ma0053P22,  
DEFINITION genomic survey sequence.  
ACCESSION BZ537360.1 GI:27085796  
VERSION BZ537360.1  
KEYWORDS GSS.  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 631)  
AUTHORS WhiteIaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,  
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
Clek,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
CONSORTIUM for Maize Genomics  
Unpublished (2002)  
COMMENT Contact: Cathy WhiteIaw  
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whiteIaw@tigr.org  
Seq primer: TF  
Class: sheared ends.  
FEATURES  
source  
1. 631  
Location/Qualifiers  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZM58Ma0053P22"  
/note="Vector: pBCK-; Site\_1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

ORIGIN  
Query Match 8.6%; Score 178; DB 8; Length 631;  
Best Local Similarity 61.8%; Pred. No. 1.4e-20;  
Matches 283; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

Qy 1559 CCGGCGGAGAGAGACCGGCGCGAGAACTGATGATCGTGAACCTGTCGCGACGACC 1618

Db 15 CCAACGAGCGAAGAAAGACCGAGAACTGATGATGTCGACCTTTAAGAAACATC 74

Qy 1619 TCAACAGCGTTCGCGGATGCGCTCGTCCACAGTGCCTTCCTTCGAGGTGAGACT 1678

Db 75 TCGGCAAGGTCTCGAGCCGAGGAGCTGATGTCACCGCTCATGAGAGTTGAATCAT 134

Qy 1679 ACCGCGCCGTCGACCAAGCTGATGTCGACATCGGCGGAGCGCTGCGCGCCGCGACGCA 1738

Db 135 ACAGAAAGTGTTCACACCATGATGATGATCGCTGCGGAAACAAAGAACGATCTAAGCC 194

Qy 1739 CCGCGCGCTCGTACGCGCGCTTCCCGCGCGCTTCATGACGCGCGCCCAAGAGC 1798

Db 195 CCGTGTATGTGTGAAGCGCGCTTCCAGCGGTTGATGACCGCGGCTCCAAAGTCA 254

Qy 1799 GCAACATGAGATCATGACCGCGCTGAGAGAGCCCGCGCGCTCTACTCCGCGCGC 1858

Db 255 GATCAATGAGATCTTGAATCACTTGAGACAAAGCCAGAGAGATGTAATCAGATCGA 314

Qy 1859 TCGGATGTTTCGCGCTCAGCGCGCGCCGACCTCAGATGTCATCCGACCATCTGTC 1918

Db 315 TCGGATGTTTCGCTGTAACAACACATTCGACCTGAAATTGTGATCGAATCTGTATCC 374

Qy 1919 TGGCCGACGCGCAGCGAGAGTTTCGCGCTCGCGCGCGAGATGTCCTCTCCGACGAG 1978

Db 375 TGCACGACGAGGAAGCTCGGTCGCGCGCGAGTGTGAGCGCTGTGAGCGCCAG 434

Qy 1979 AGGAGAGTTTACCGAGACCGTGTAAAGCCCGCC 2016

Db 435 AAGCAGAGTACGACGAATGCTCTGAAAGCGAAAGCC 472

RESULT 8  
CG170785 810 bp DNA linear GSS 21-AUG-2003  
LOCUS PU1R58TD\_ZM\_0.6\_1.0\_KB\_Zea\_mays\_genomic\_clone\_ZM58Ta0592120,  
DEFINITION genomic survey sequence.  
ACCESSION CG170785  
VERSION CG170785.1 GI:34061583  
KEYWORDS GSS.  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 810)  
AUTHORS WhiteIaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,  
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and  
Bennetzen,J.  
Maize Genomics Consortium  
Unpublished (2003)  
COMMENT Other\_GSSs: PU1R58TB  
Contact: Cathy WhiteIaw  
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whiteIaw@tigr.org  
Seq primer: TF  
Class: sheared ends.  
FEATURES  
source  
1. 810  
Location/Qualifiers  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZM58Ta0592120"  
/note="Vector: pCR4-TOPO; Site\_1: EcoRI; 0.6-1.0 kb high  
Cor selected genomic DNA library"

ORIGIN  
Query Match 8.6%; Score 178; DB 9; Length 810;  
Best Local Similarity 61.8%; Pred. No. 1.4e-20;  
Matches 283; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

Qy 1559 CCGGCGGAGAGAGACCGGCGCGAGAACTGATGATCGTGAACCTGTCGCGACGACC 1618

Db 663 CCAACGAGCGAAGAAAGACCGAGAACTGATGATGTCGACCTTTAAGAAACATC 604

Qy 1619 TCAACAGCGTTCGCGGATGCGCTCGTCCACAGTGCCTTCCTTCGAGGTGAGACT 1678

Db 603 TCGGCAAGGTCTCGAGCCGAGGAGCTGATGTCACCGCTCATGAGAGTTGAATCAT 544

Qy 1679 ACCGCGCCGTCGACCAAGCTGATGTCGACATCGGCGGAGCGCTGCGCGCCGCGACGCA 1738

Db 543 ACAGAAAGTGTTCACACCATGATGATGATCGCTGCGGAAACAAAGAACGATCTAAGCC 484



QY 1739 CCGCCGCGCTTCGTAACGCGCCGCTTCCCGCGGCTTCATACCGCGCGCCCAAGAGC 1798  
Db 483 CCGTGTATGTGTGAAGAGCGGCTTTCAGCGGTTGATGACCGGGGCTCCAAAGTCA 424  
QY 1799 GCACCATGAGATCATGACCCGCTGGAGGAAGCCCGCGGCTTACTTCGGGGCGC 1858  
Db 423 GATCAATGGAGATCTTGATTCATTGAGACAAGCCCAAGAGAGTGATCTTCAGGATCGA 364  
QY 1859 TCGGATGATTCGCGCTCAGCGGCGCGCCGACCTCAGCATGTCATCCGACCATGATGTC 1918  
Db 363 TCGGATTCCTCTGTGATACACACATTCGACCTGAAATTTGATTCGAACTGTTATCC 304  
QY 1919 TGGCCGACGCGCAAGCGGAGTTCCGCGCTCGCGGGGATGTCGCTCTCCGACGAG 1978  
Db 303 TGCACGACGGAAGAGCTCGGTGCGGGGCGGGTGGGGGATGTAAGCTGTGACGCGCAG 244  
QY 1979 AGGAGAGTTACCGAGACCGGTGTAAAGCCCGCGCC 2016  
Db 243 AAGCAGATGACGAGAAATGCTGTGAAAGCGAAAGCC 206

RESULT 9  
CF598037 579 bp mRNA linear EST 26-SEP-2003  
LOCUS NC053906.12 y1 NC-LIV tachyzoite cDNA library Neospora caninum  
DEFINITION CF598037  
ACCESSION CF598037  
VERSION CF598037.1 GI:36356095  
KEYWORDS EST.  
SOURCE Neospora caninum  
ORGANISM Neospora caninum  
Bukaryotes; Alveolata; Apicomplexa; Coccidia; Eimeriida;  
Sarcocystidae; Neospora.  
REFERENCE 1 (bases 1 to 579)  
AUTHORS Cole, R., Fogarty, S., Tang, K., Howe, D. K., Sibley, L. D., Clifton, S.,  
Warr, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Theising, B.,  
Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Ronko, I.,  
Teagarden, J. R., Fedele, M., Belaygorod, L., Franklin, C.,  
Carr, L. M., Grow, A., Maguire, L., Wadkins, J., Ritchey, J., Waterston, R.  
and Wilson, R.  
TITILE USDA-WashU Neospora EST Project  
JOURNAL Unpublished (2000)  
COMMENT Contact: Sandy Clifton, Ph.D. - Neospora  
USDA-WashU Neospora EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
Contact David Sibley (toxoe@borcim.wustl.edu) for further  
information relating to organism, libraries, or clone availability.  
Seq primer: -40RP from Gibco.  
Location/Qualifiers  
1. 579  
/organism="Neospora caninum"  
/mol\_type="mRNA"  
/db\_xref="taxon:29176"  
/dev\_stage="tachyzoite"  
/lab\_host="Electroten Blue cells (Stratagene)"  
/clone\_lib="NC-LIV Tachyzoite cDNA library"  
/note="Vector: pBluescript II SK+vector type: plasmid;  
Site 1: EcoRI; Site 2: XhoI; The cDNA library was  
constructed by Kelian Tang, and Robert Cole at Washington  
University. cDNA was synthesized from Poly(A)+ mRNA using  
an oligo-d(T) primer containing a XhoI site. Following  
second strand synthesis, EcoRI adapters were ligated to  
the cDNA, and products were size-selected on sephacryl  
S500. The cDNA were directionally cloned into the  
EcoRI/XhoI prepared pBluescript II SK+ vector, and  
electroporated into Electroten Blue cells (Stratagene).  
The library may contain a small percentage of host or  
bacterial contaminants."

ORIGIN  
Query Match 8.5%; Score 174.2; DB 7; Length 579;  
Best Local Similarity 61.5%; Pred. No. 6 4e-20;  
Matches 297; Conservative 0; Mismatches 183; Indels 3; Gaps 1;

QY 1444 CCGGAGCGGTTCTCTCAAGATCGGCGCGGCGGCTGAGTCCAAAGCCATCAAGGGG 1503  
Db 76 CCGGAACGCTTTCTAATGAAGATGAGCGGTTGGCTGAGATCGAAACCATCAAGGGT 135  
QY 1504 ACCCG 1563  
Db 136 ACAGGCG 195  
QY 1564 CCGGAGAAAGACCGGCGCGGACGAGACCTGATGATCGTCACTGCTCGCAACGACCTCAAC 1623  
Db 196 AAGCAAAAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 255  
QY 1624 AGCGTTCGCGGATCGGCTCGGTCACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1683  
Db 256 CGGTCTGTATTCAGGAACCGTGACGCTTCTGAGTTGATGATGATGATGATGATGATG 315  
QY 1684 CCGGTGACACGCTGTGTGACACATCCGGGACCGCTGCGG---CCGGCACACGACCC 1740  
Db 316 ACAGTGCATCAACTGTGTGACCAATCCGAGCGAAGCTGAGGAAACCGAAGTGAATT 375  
QY 1741 GCGGCTGCGTACG 1800  
Db 376 CTTGACGCTGTATATGCGCATTTCCAGAGGCTGATACAGCGGCTCCAGAGAGCGG 435  
QY 1801 ACCATGAGATCATGACCGCGCTGAGAGAGGCGCGCGCGCGCGCGCGCGCGCGCG 1860  
Db 436 GCAATGAAATCATGAAAGGCTGAGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 495  
QY 1861 GGATGTTGCCCTCAGCG 1920  
Db 496 GCGTTCCTGTGCGTACAGGTGAAGCAAGTAAACATGTCATTCGAACGTCGCTTT 555  
QY 1921 GCC 1923  
Db 556 ACC 558

RESULT 10  
B2892725/c 660 bp DNA linear GSS 30-JUL-2003  
LOCUS B2892725/c  
DEFINITION Hm7\_0152.x1 059.ab1 Hm pUC18 library Haloarcula marismortui ATCC  
43049 genomic 5', genomic survey sequence.  
ACCESSION B2892725  
VERSION B2892725.1 GI:3343358  
KEYWORDS GSS.  
SOURCE Haloarcula marismortui ATCC 43049  
ORGANISM Haloarcula marismortui ATCC 43049  
Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloarcula.  
REFERENCE 1 (bases 1 to 660)  
AUTHORS Goo, Y., Roach, J., Glusman, G., Baliga, N. S., Deutsch, K., Pan, M.,  
Dasarma, S., Ng, W. V. and Hood, L.  
TITILE Low-pass Sequencing for Microbial Comparative Genomics  
JOURNAL Unpublished (2003)  
COMMENT Contact: Goo Y  
Institute for Systems Biology  
1441 North 34th Street, Seattle, WA 98103, USA  
Tel: 206 732 1412  
Fax: 206 732 1299  
Email: ygoo@systemsbiology.org  
Seq primer: M13 Forward  
Class: shotgun.  
Location/Qualifiers  
1. 660  
/organism="Haloarcula marismortui ATCC 43049"  
/mol\_type="genomic DNA"  
/strain="ATCC 43049"



/db\_xref="taxon:272569"  
/clone\_1lb="Hm pUC18 Library"  
/note="Vector: pUC18; Site 1: SmaI; A shotgun library was  
constructed from Haloarcula marismortui genomic DNA using  
pUC18/SmaI/BAP plasmid"

## ORIGIN

Query Match 8.3%; Score 171.8; DB 8; Length 660;  
Best Local Similarity 56.1%; Pred. No. 1.6e-19;  
Matches 345; Conservative 0; Mismatches 267; Indels 3; Gaps 1;

QY 1239 CGACAAAGAGCGCTTACCTCAAGCGCATGAGAGTGCCTCAAGAGATCCGCAACGCGCA 1298  
DB 619 CGGGAAGACAGCGTTCCCGACCGCGTCCGCAAAATAAACAGTACGTCAAGGAGCGCGCA 560  
QY 1299 GTGTGATGAGATCTGCTGACCAACATGTCACCGCGCCGACCGAGCGACGCGCTGCC 1358  
DB 559 CAGGTTCCAAACAAACCTCTCACACCGCTGACCGCCCGCGCGCTCCACCCAGTCA 500  
QY 1359 GCTTACTCCGCGCGCGCATCAGCCCGTCCGCTAGAGGCGCTGCTCGAGTTCC 1418  
DB 499 CACCTTTCAGCTGTCTGTCGGGTGAATCCGCTCCGACTCAGCCCTGCTGAGTTCC 440  
QY 1419 CGAATCTCGGTGCTGACCGCTTCGCCGAGCGGTTCTCAAGATCGGCGCCGACGCGG 1478  
DB 439 CGGCGTGAACCTTGTACCGCGACCGCGAGCTGTTCT---GGACCTCATGCGCGCA 383  
QY 1479 GCTGAGTCCAAAGCCCATCAAGGGAGCCCGCCCGGGGCGCAACCGCGAGAGACGA 1538  
DB 382 GTTGCTCAGCAGAGCGAATGCCCGMACCGCGCGCGAGCGCGCTGCCAAGACGA 323  
QY 1539 GCGGCTCCGCGCGACCTGCGCGCGCGGAGAGGACCGGGCGGAGAACCTGATGATCGT 1598  
DB 322 GGAACCTTGAAGTGAACCTTCTGTAACGACGAGAGAGAGCGGCTGACACCGCATGCTGT 263  
QY 1599 CGAATCTGTCGCGCAACGACCTCAACAGCGTCTGCGCGATCGGCTCCGTCACGTCGCCG 1658  
DB 262 CGAATCTGAAACGCAATGACCTCGGCAAGTCAAGGATATGGTCCGTGACGTCGCCGA 203  
QY 1659 GCTCTTGAAGTGAAGACCTAAGCGCGCGGTGACACGACTGCTGTCACATCCGAGGAG 1718  
DB 202 GTACCGCGCGGTGACCGGTATTCGAAAGAAAGCACTGTTTCCCTCATCAAGGAGGA 143  
QY 1719 GCTGCGCGCGCGGACACGACCGCGCTGCTGACGCGCGCTTCCCGCGCGCTTCAT 1778  
DB 142 ATTACGGAGACGGGTGACATCGCGACGCGGTGCGACGCGTTCCTGCTGAGACAT 83  
QY 1779 GACCGCGCGCGCAAGAGCGCACATGAGATCATCGACCGCTGAGAGAGAGGCGCGG 1838  
DB 82 CACCGCGCGCGCAAGCGCGGACGATGAGATTATGACGAGTGAAGCGGACCGGACG 23  
QY 1839 GGGCGTCTACTCGCG 1853  
DB 22 GGGGCGCTTACACCGG 8

RESULT 11  
COORD4641/c 894 bp mRNA linear EST 10-JUN-2004  
LOCUS CO024641  
DEFINITION Coccidioides posadasii spherule cDNA library, 0.5 to 5.3  
ACCESSION CO024641  
VERSION CO024641.1 GI:48551893  
KEYWORDS EST.  
SOURCE Coccidioides posadasii  
ORGANISM Coccidioides posadasii  
Eukaryote; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Omygenales; mitosporic Omygenales; Coccidioides.  
REFERENCE 1 (bases 1 to 894)  
AUTHORS Gardner,M.J. and Cole,G.T.  
TITLE Analysis of gene expression in Coccidioides posadasii mycelia and  
spherules via expressed sequence tags  
JOURNAL Unpublished (2003)

## COMMENT

Other ESTs: EST803026  
Contact: Gardner MJ  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301 838 3519  
Fax: 301 838 0208  
Email: gardner@igr.org.

FEATURES  
source Location/Qualifiers

1..894  
/organism="Coccidioides posadasii"  
/mol\_type="mRNA"  
/strain="C735"  
/db\_xref="taxon:199306"  
/clone="CIFA227"  
/dev\_stage="spherules"  
/lab\_host="E. coli DH10B, T1 phage resistant"  
/clone\_1lb="Coccidioides posadasii spherule cDNA library,  
0.5 to 5.3 kb"  
/note="Vector: pExpress 1; Site 1: Not I; Site 2: Eco RV;  
Coccidioides posadasii spherule cDNA library, 0.5 to 5.3  
kb"

## ORIGIN

Query Match 8.1%; Score 167.8; DB 7; Length 894;  
Best Local Similarity 52.6%; Pred. No. 7.6e-19;  
Matches 442; Conservative 0; Mismatches 387; Indels 12; Gaps 3;

QY 1231 GCGCGCACGACAAAGAGCGCTTACCTCAAGCGCATGACGAGTGCCTCAAGAGATCCG 1290  
DB 841 GCCCTGAGACGACGAACATCTGTACACGAGAGATCGGAGCGGACAGAAATGCTTAAG 782  
QY 1291 AAGGAGATCTGATGAGATCTGCTGACCAACATGATGATCGGCGCG---ACGAGGCG 1347  
DB 781 GAGGAGGAGCGGTATGAGTGAACATACGACCAAAATTAAGGCATGCTGTCGATGTA 722  
QY 1348 ACGGCGCTCGCTTACTCTCCGCGTCCGCGCATCAGCGCGCGCTCCGTAAGCGCGCTG 1407  
DB 721 GACCCATAGCATTAATATCTCTCCCTCGAAGAACGTAACCGCCCTTACTAGCGTAT 662  
QY 1408 CTCGAGTTCC---CCGAATGTCGTGTGAGGCGCTCGCGCGGTTCTTCACG 1461  
DB 661 ATGAGCTTCGCTGTCACAGACACCAATATCTCTTCGCGCCGAGAGTTATATATCC 602  
QY 1462 ATGAGGCGCGAGCGGCGCTGAGTCCAGGTCATCAAGGAGACCGCGCGCGCGCGG 1521  
DB 601 ATGATGCTGATGCGCTGCTGTAATGAACCAATTAAGGAAACCTTCCCTGACGCCA 542  
QY 1522 ACCGCGAGAGAGAGAGCGGCGCTCCGCGCGACCTGCGCGCGCGGAGAGAGCGGCG 1581  
DB 541 GACAAAG 482  
QY 1582 GAGAACCTGATGATCTGTCAGCTGTGCGCAACGACCTTCAACAGCTGCGGATCGG 1641  
DB 481 GAGAACTTAATGATTTGATCTTATTCGCGCGGACCTTCAACAACTTCTCATCAAA 422  
QY 1642 TCCGTCAAGTGCCTCGGCTCTTTCAGGTGAGAGACTTACGCGCGCTGCAACAGCTG 1701  
DB 421 TCATCAAGGTCCCGAATCTACCTGCTGTAAGAGTCAAGAAACATGATCATGAGCTGTC 362  
QY 1702 TCGACCATCGGGGAGAGGCTGCGCGCGGACAGACAGCGCGCGCTCGGTAAGCGCGCG 1761  
DB 361 ACAACATTTCAATCCCAATCGCGCCCAACGTCGCGCGGCTCAGGTTCTCAGAGCATG 302  
QY 1762 TTCGCGCGGCGCTCATGACCGCGCGCGCAAGAGCGCACCATGAGATCATGACCGC 1821  
DB 301 TTCACACAGATTCATGACGCGCGCTCCCAACTCAGAGCGGTGCAAAATCTTCGACGCG 242  
QY 1822 CTGAGAGAGGCGCC---CGGCGCTTACTCTCGGCGCGCTGAGATGTTCCGCTTCAC 1878  
DB 241 CTGGAAGACACCGACGAGCGGCAATATCTCCGCGAGTCTGGGCTATGATGCGCGAC 182  
QY 1879 GGGCGCGCGACCTGACATCTGATCCGACACCATGCTCTGCGGACGCGGCGGAG 1938



Db 181 GGACCGTCGATGATGCGTGTATCCGATGATGTAAGTAAAGACGGAAGCAGCTTGAG 122  
QY 1939 TTCCGCGCTCGCGCGGCGGATCGTGTCCCTCCGACCGAGGAGGAGTTCAACGAGACC 1998  
Db 121 CTAGTGTCTGAGAGGCGGATTACTTGCTCAGCGAGCGGAGAAAGATGGATGAGTGC 62  
QY 1999 GTGTAAAGCCCGCCGATGTCACCCCTCGACGCGAGCGCGTGGCGGCGCCCGA 2058  
Db 61 ATGGTGAAGCGATGCTGTGCTACGCGCTCGCTTGAGATCGGCGCAGATGCTGGA 2  
QY 2059 T 2059  
Db 1 T 1

RESULT 12  
LOCUS BZ896109 757 bp DNA linear GSS 30-JUL-2003  
DEFINITION NARP6\_0132 Na pUC18 library Natrialba asiatica genomic 5', genomic survey sequence.  
ACCESSION BZ896109  
VERSION BZ896109.1 GI:33346585  
KEYWORDS GSS.  
SOURCE Natrialba asiatica  
ORGANISM Natrialba asiatica  
Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Natrialba.  
REFERENCE 1 (bases 1 to 757)  
AUTHORS Gao,Y., Koach,U., Glusman,G., Baliga,N.S., Deutsch,K., Pan,M., Dassaerma,S., Ng,W.V. and Hood,L.  
TITLE Low-pass Sequencing for Microbial Comparative Genomics  
JOURNAL Unpublished (2003)  
COMMENT Contact: Gao Y  
Institute for Systems Biology  
1441 North 34th Street, Seattle, WA 98103, USA  
Tel: 206 732 1412  
Fax: 206 732 1299  
Email: YGao@systemsbiology.org  
Seq primer: M13 Forward  
Class: shotgun.  
FEATURES  
source 1..757  
Location/Qualifiers  
/organism="Natrialba asiatica"  
/mol\_type="genomic DNA"  
/strain="ATCC 700177"  
/db\_xref="taxon:64602"  
/clone\_lib="Na pUC18 Library"  
/note="Vector: pUC18, Site\_1: SmaI; A shotgun library was constructed from Natrialba asiatica genomic DNA using pUC18/SmaI/BAP plasmid"

ORIGIN  
Query Match 8.1%; Score 167.2; DB 8; Length 757;  
Best Local Similarity 57.9%; Pred. No. 9.8e-19;  
Matches 334; Conservative 0; Mismatches 229; Indels 4; Gaps 2;

Db 1361 TCTACTCCGCGCGCGCCATCAGCCCGCTCCGTCAGCGCCCTGCTGATGTTCCCG 1420  
QY 90 TCTACGAGCGATGCGGAGATCAACCTCGCGCTGATCTGCTGAGACGACG 149  
Db 1421 AACTGTGATGCTGAGAGCGCTCGCCGAGCGGTTCTTCAGATGCGCGAGCGCGCG 1480  
QY 150 ACCTGACCGTCTGTGGTGCAGAGCCCGAAGACGCTGCTATGGTGGCGGCGC---GAGAG 206  
Db 1481 TCGATCCAAGCCCATCAAGGGAACCGCCCGGCGGCGACCGCGAGAGAGAGCAGC 1540  
QY 207 TCATGCGGAATCCGATCGCGGCACCTGCGACCGCGGAGAGAGCGCCGTCGAGGACCGC 266  
Db 1541 GGCTCCGCGCGCACTGCGCGCGCGGAGAAAGACCGGCGCGAGAACTGATATGCTCG 1600  
QY 267 GACTGGCGGCGAATCTGCGCAGACGGGAAAGAGCGCGCGCAACACCATGCTGTGCTG 326  
Db 1601 ACTGTGTCGCAAGCACTCAAGCGCTGCGCGATCGGCTCGTCCATGTCGCCCGCGC 1660

Db 327 ACCTCGCGGAACGATGTAACGCCGCTCTCGCGCGCGGCTCCGTCGCTGACGAGT 386  
QY 1661 TCTTCGAGGTGAGACCTACCGGCGCCCTGACACACTGTGTGACACATCCGGGGACGCG 1720  
Db 387 TCATGAACGTGCTGAAGTACACGCACTGACACACATGACGAGCGGTGACGAGGAGAT 446  
QY 1721 TCGCGCCCGGACACGACCGCGCGCTGCTGATGCGCGCGCTTCCCGCGGCGCTCATGA 1780  
Db 447 TGCGCGCGGACCGCCGACCGGCTTGAACGCGAGCGCGGCGCTGTTCCCGCGGAGCGCT 506  
QY 1781 CCGGCGCGCCCAAGAAAGCGCACATGAGATATGACCGGCTGAGAGAGGCCCGCGG 1840  
Db 507 CGGCGCGCGCAAGATCCGCGGATGAGCTCATGCACTGACGACCGAGCCGCGCG 566  
QY 1841 GCGTACTCCGCGGCGC--TCGGATGTTCCGCTCAAGCGCGCGCGGCACTCAAGATC 1899  
Db 567 GGCTGTACGCGGAGGCGCTCGCTACTACTCTTGAGACGCGCGAGATTTCCGATC 626  
QY 1900 GTCATCCGACCATCGTGTGCGCGACGCGCGAGCGG 1936  
Db 627 GTCATTCGACGCGCACGCTGGAAGCGAGCGCGCGG 663

RESULT 13  
LOCUS CO028609/c 750 bp mRNA linear EST 10-JUN-2004  
DEFINITION EST806993 Coccidioides posadasii spherule cDNA library, 0.5 to 5.3 kb Coccidioides posadasii cDNA clone CIPAS82 3' end, mRNA sequence.  
ACCESSION CO028609  
VERSION CO028609.1 GI:4859837  
KEYWORDS EST.  
SOURCE Coccidioides posadasii  
ORGANISM Coccidioides posadasii  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Onygenales; mitosporic Onygenales; Coccidioides.  
REFERENCE 1 (bases 1 to 750)  
AUTHORS Gardner,M.J. and Cole,G.T.  
TITLE Analysis of gene expression in Coccidioides posadasii mycelia and spherules via expressed sequence tags  
JOURNAL Unpublished (2003)  
COMMENT Other ESTs: EST806994  
Contact: Gardner MJ  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301 838 3519  
Fax: 301 838 0208  
Email: gardner@tigr.org.  
FEATURES  
source 1..750  
Location/Qualifiers  
/organism="Coccidioides posadasii"  
/mol\_type="mRNA"  
/strain="C735"  
/db\_xref="taxon:199306"  
/clone\_lib="CIPAS82"  
/dev\_stage="spherules"  
/lab\_host="E. coli DH10B, T1 phage resistant"  
/clone\_lib="Coccidioides posadasii spherule cDNA library, 0.5 to 5.3 kb"  
/note="Vector: pEXpress 1; Site\_1: Not I; Site\_2: Eco RV; Coccidioides posadasii spherule cDNA library, 0.5 to 5.3 kb"

ORIGIN  
Query Match 8.0%; Score 164.8; DB 7; Length 750;  
Best Local Similarity 55.1%; Pred. No. 2.5e-18;  
Matches 344; Conservative 0; Mismatches 277; Indels 3; Gaps 1;

QY 1439 CCTCGCGGAGGCGTCTTCAGATGCGCGCGCGCGCGCTGACATCCAGCCCATCA 1498  
Db 624 CGTCCGCCGAGAGGTTTATATCATGATGCTGATGCGCTGCTGAATGAACCAATGA 565  
QY 1499 AGGGAGCCGCGCGCGGCGGCGACCGCGAGAGAGACGAGCGGCTCCGCGCGCACTTG 1558



```
Db 564 AGGAAACCTTGGCCGTCAGCCAGACAAAGAAAGCGAGGAGAAAAAGTCAAGTTAG 505
Qy 1559 CCGGCGGAGAAAGACCGGGCCGAGAACTGATGTCGTGACCTGTGTCCGACGACC 1618
Db 504 CGACGATGTCAGAGAGCTCGCCGAGAACTTAATGATTTGATCTTATCCGCGCGAGCC 445
Qy 1619 TCAACAGCGCTCGCGGATCGGCTCGCTCCAGTGGCCCGGCTCTTCAGAGTGAACCT 1678
Db 444 TCCACACATTTTCCATCAAAATCAATCAAGTCCGAAACTACTCTCAGTTGAAGCT 385
Qy 1679 AGCGCCCGTCGACACGAGTGTGTGACCATCCGAGGAGCGGTGCGGCCGAGCAGCA 1738
Db 384 AGAAACAGTACATGACTGTGTGACCAACATTCATCCCATCTGCCCCCAAGTGGCG 325
Qy 1739 CCGCGCGCTGCGTACCGCGCCGCTTCCCGCGGCTTCATGACGCGCGCCCAAGAGC 1798
Db 324 GCGTCCAGGTTCTCGACGCGATGCTTCCACAGAGATCATGACGGGCGCTCCAAACTCA 265
Qy 1799 GCACCATGAGATCATCGACCGGCTGAGAGAGAGGCCCC---CGGGCGCTTACTCCGGGG 1855
Db 264 GAGCCGTGAAATCTTCGACGCGCTGAGAGAGACCGCAGCGCGGATATCTCCGCGCA 205
Qy 1856 CGCTCGGATGTTTCCGCTTCAGCGGCGCGCCGACCTCAGCATGTCATCCGACCATCG 1915
Db 204 GTCTGGGCTATGATATGCGCGCAGCGCAGCCTGATCAGTCGGTGTATTCGTAACATTTG 145
Qy 1916 TGTGCGCCGACGCGCAGCGGAGATTGCGGCTCGCGGGCGGATGCTGTCCTCTCCGACC 1975
Db 144 TAAAGTACGGAAGAGCTTGAGCTGAGTGTGAGAGGGCGGATTAATCTGCTCAGAGAG 85
Qy 1976 AGGAGAGAGATTACCGAGACCGTGTAAAGCCCGCGCATGTCACCGCCCTCGAGC 2035
Db 84 CGAGAAAGATGGATGAGTGTATGTGTAAGGCGAATGTGTGTACGCGCTGCTCTTA 25
Qy 2036 GCAGCGCCGTGCGGCGCGCCGAT 2059
Db 24 GGGAAATCGGCGCGAATGCTGTGAT 1
```

```
RESULT 14
COO30342/c 818 bp mRNA linear EST 10-JUN-2004
LOCUS EST080876 Coccidioides posadasii spherule cDNA library, 0.5 to 5.3
DEFINITION kb Coccidioides posadasii cDNA clone C1FB368 3' end, mRNA sequence.
ACCESSION COO30342
VERSION COO30342.1 GI:48563318
KEYWORDS EST.
SOURCE Coccidioides posadasii
ORGANISM Coccidioides posadasii
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; Mitosporic Onygenales; Coccidioides.
1 (bases 1 to 818)
Gardner, M.J., and Cole, G.T.
Analysis of gene expression in Coccidioides posadasii mycelia and
spherules via expressed sequence tags
Unpublished (2003)
Other ESTs: EST808727
Contact: Gardner MJ
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@tigr.org.
Location/Qualifiers
1..818
/organism="Coccidioides posadasii"
/mol_type="mRNA"
/strain="C735"
/db_xref="taxon:199306"
/clone="C1FB368"
/dev_stage="spherules"
/lab_host="E. coli DH10B, T1 phage resistant"
```

FEATURES

source

## ORIGIN

Query Match 8.0%; Score 164.8; DB 7; Length 818;  
Best Local Similarity 55.1%; Pred. No. 2.5e-18;  
Matches 344; Conservative 0; Mismatches 277; Indels 3; Gaps 1;

```
Qy 1439 CCTGGCCGAGCGGTTCTCAGATCGGCGCGAGCGGCGCGCTCGAGTCCAAAGCCATCA 1498
Db 624 CGTGCGCCGAGAGGTTTATATCATATGATGATGATGATGATGATGATGATGATGATG 565
Qy 1499 AGGAGACCCGCGCGCGGCGGACCGCGAGAGAGACGAGCGGCTCCGCGCGACCTGG 1558
Db 564 AGGAAACCTTGGCCGTCAGCCAGACAAAGAAAGACGAGGAGAAAAAGTCAAGTTAG 505
Qy 1559 CCGGCGGAGAAAGACCGGGCCGAGAACTGATGATGATGATGATGATGATGATGATGATG 1618
Db 504 CGACGATGTCAGAGAGCTCGCCGAGAACTTAATGATTTGATCTTATCCGCGCGAGCC 445
Qy 1619 TCAACAGCGCTCGCGGATCGGCTCGCTCCAGTGGCCCGGCTCTTCAGAGTGAACCT 1678
Db 444 TCCACACATTTTCCATCAAAATCAATCAAGTCCGAAACTACTCTCAGTTGAAGCT 385
Qy 1679 AGCGCCCGTCGACACGAGTGTGTGACCATCCGAGGAGCGGTGCGGCCGAGCAGCA 1738
Db 384 AGAAACAGTACATGACTGTGTGACCAACATTCATCCCATCTGCCCCCAAGTGGCG 325
Qy 1739 CCGCGCGCTGCGTACCGCGCCGCTTCCCGGCGCTTCATGACCGCGCGCCCAAGAGC 1798
Db 324 GCGTCCAGGTTCTCGACGCGATGCTTCCACAGAGATCATGACGGGCGCTCCAAACTCA 265
Qy 1799 GCACCATGAGATCATCGACCGGCTGAGAGAGAGGCCCC---CGGGCGCTTACTCCGGGG 1855
Db 264 GAGCCGTGAAATCTTCGACGCGCTGAGAGAGACCGCAGCGCGGATATCTCCGCGCA 205
Qy 1856 CGCTCGGATGTTTCCGCTTCAGCGGCGCGCCGACCTCAGCATGTCATCCGACCATCG 1915
Db 204 GTCTGGGCTATGATATGCGCGCAGCGCAGCCTGATCAGTCGGTGTATTCGTAACATTTG 145
Qy 1916 TGTGCGCCGACGCGCAGCGGAGATTGCGGCTCGCGGGCGGATGCTGTCCTCTCCGACC 1975
Db 144 TAAAGTACGGAAGAGCTTGAGCTGAGTGTGAGAGGGCGGATTAATCTGCTCAGAGAG 85
Qy 1976 AGGAGAGAGATTACCGAGACCGTGTAAAGCCCGCGCATGTCACCGCCCTCGAGC 2035
Db 84 CGAGAAAGATGGATGAGTGTATGTGTAAGGCGAATGTGTGTACGCGCTGCTCTTA 25
Qy 2036 GCAGCGCCGTGCGGCGCGCCGAT 2059
Db 24 GGGAAATCGGCGCGAATGCTGTGAT 1
```

```
RESULT 15
COO34510/c 820 bp mRNA linear EST 10-JUN-2004
LOCUS EST12894 Coccidioides posadasii spherule cDNA library, 0.5 to 5.3
DEFINITION kb Coccidioides posadasii cDNA clone C1FB103 3' end, mRNA sequence.
ACCESSION COO34510
VERSION COO34510.1 GI:48571664
KEYWORDS EST.
SOURCE Coccidioides posadasii
ORGANISM Coccidioides posadasii
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; Mitosporic Onygenales; Coccidioides.
1 (bases 1 to 820)
Gardner, M.J., and Cole, G.T.
Analysis of gene expression in Coccidioides posadasii mycelia and
spherules via expressed sequence tags
Unpublished (2003)
```



## COMMENT

Other\_ESTs: EST812895  
Contact: Gardner MJ  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301 838 3519  
Fax: 301 838 0208  
Email: gardner@igr.org.

## FEATURES

source

Location/Qualifiers  
1..820  
/organism="Coccidioides posadasii"  
/mol\_type="mRNA"  
/strain="C735"  
/db\_xref="taxon:199306"  
/clone="C1FBU13"  
/dev\_stage="spherules"  
/lab\_host="E. coli DH10B, T1 phage resistant"  
/clone\_lib="Coccidioides posadasii spherule cDNA library,  
0.5 to 5.3 Kb"  
/note="Vector: pExpress 1; Site\_1: Not I; Site\_2: Eco RV;  
Coccidioides posadasii spherule cDNA library, 0.5 to 5.3  
Kb "

## ORIGIN

Query Match 8.0%; Score 164.8; DB 7; Length 820;  
Best Local Similarity 55.1%; Pred. No. 2.5e-18;  
Matches 344; Conservative 0; Mismatches 277; Indels 3; Gaps 1;

```
QY 1439 CCTGCGCCGAGGGTTCTCAGATCGGCGCGAGCGGCGCTCGAGTCCAGCCCATCA 1498
DB 624 CGTCGCCGAGAGGTTATATCATAGATCTGATGCGCTGTAATGAATGAACAAATA 565
QY 1499 AGGGGAGCCGCGCCCGGCGCGACCGCGAGAGGAGAGAGAGCGGCTCCGCGCCGACCTGG 1558
DB 564 AGGGAACCTTGGCCGTACGCCAGACAAAGAGAGAGAGCGGAGAAAAGTCACTTAG 505
QY 1559 CCGGCGGAGAAAGACCGGCGCGAGACCTGATGATGCTGACCTGTGCGCAGAC 1618
DB 504 CGAGGATGTCAGAGGCTCGCCGAGAACTTAATGATTCATCTTATCCGCGCGAC 445
QY 1619 TCACAGCGCTGCGGCGATCGGCTCCGTCACGTCGCCGCGCTTCGAGGTGAGACCT 1678
DB 444 TCACAAACATTTTCTCAATAAATCAATCAAGGTCCGAAACTATCTCAGTTGAAAGCT 385
QY 1679 ACCGCGCGGTGACACGAGTGTGACCATCCGGGAGCGGCTGCGCGCGGACCGACA 1738
DB 384 ACGAACAGTACATCACTGCTGTCACAACATTCATCCCATCGCCCCCAAGTCGGCG 325
QY 1739 CCGCGCGCTGCGTACGCGCGCTTCCCGCGGCTTCATGACCGCGCGGCCAAGAAGC 1798
DB 324 GCGTCAGGTTTCGAGCGATGCTCCACACAGATTCATGACGCGCGCTCCCAACTCA 265
QY 1799 GCACATGAGATCATGACCGCTGAGAGAGGCCCC--CGGGCGTCTACTCCGGGG 1855
DB 264 GAGCCGTCAAAATCTTGAGCGGCTTGAGAGACACCGCGCGGATATCTCTCGGCA 205
QY 1856 CGCTCGATGTTGCGCTTCAGCGGCGCGCGACCTCAGCATGTCATCCGACCATCG 1915
DB 204 GTCTGGGCTATGATGCGCGAGCGGACCGTCGATCAGTCGTTATCCGACATG 145
QY 1916 TGCTGGCCGACGCGCAGCGGAGTTGCGGCTCGCGGGGCGATGTCCTCTCCGACC 1975
DB 144 TAAAGTACGGAAGCAGCTTAGCTGAGAGGCGGCGATTACTTGGCTCAGCAGG 85
QY 1976 AGGAGGAGGATTCACCGAGACCGTGTAAAGGCCCGCGCATGTCACCGCCCTGACG 2035
DB 84 CGAGAAAGATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 25
QY 2036 GCAGCGCGCTGCGCGGCGCGCAT 2059
DB 24 GGGATTCGGCGCCAGATGCTGAT 1
```

Job time : 7804.14 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comugen Ltd.

## OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 05:31:11 ; Search time 10157.2 Seconds  
(without alignments)  
3272.568 Million cell updates/sec

Title: US-10-089-514-2

Perfect score: 3573  
Sequence: 1 MRTLIDNVDSTHNLFOYI.....VVRAMVTLADGSAVAGAR 686

## Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seg length: 0  
Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh  
-Q=/c92.1/USFTO.spool.2/US10089514/runat\_04102005\_105744\_8033/bpp.query.fasta\_1.1621  
-DB=GenEmbl -QFMT=fastap -SUFFIX=c9e -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NORM=gct -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10089514.@CCN.1.1.8782.@runat\_04102005\_105744\_8033 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hlg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3573	100.0	2061	6	BD178313 Transform
2	3573	100.0	2061	6	BD093914 Transform
3	3573	100.0	5251	1	AB116234 Streptomy
4	3127.5	87.5	2660	1	SVU21728 Streptomy

5	3127.5	87.5	14159	1	AF262220 Streptomy
6	1816	50.8	2220	6	AR198361 Sequence
7	1816	50.8	4740	1	SPU60417 Streptomy
8	1766	49.4	4607	1	STMABA Streptomy
9	1766	49.4	39314	1	SGR300302 Streptomy
10	1766	49.4	138203	1	AY310323 Streptomy
11	1635.5	44.8	110000	1	AP006618 Streptomy
12	1606	44.9	333550	1	AP003593 Streptomy
13	1600.5	44.8	29132	1	STH575648 Streptomy
14	1596.5	44.7	283300	1	AP005025 Streptomy
15	1541	43.1	300750	1	AP005217 Streptomy
16	1487	41.6	1860	6	BD163312 Novel pol
17	1487	41.6	1860	6	AX121195 Sequence
18	1487	41.6	1983	6	AX064177 Sequence
19	1487	41.6	1983	6	AX064419 Sequence
20	1487	41.6	2005	6	AX764213 Sequence
21	1487	41.6	333150	1	AP005277 Coryneb
22	1487	41.6	349459	1	BX927151 Coryneb
23	1487	41.6	349980	6	AX127146 Sequence
24	1475.5	41.3	349944	1	BX571871 Photohab
25	1475.5	41.3	349980	6	AX770909 Sequence
26	1324	37.1	3174	8	AY425708 Lycopersi
27	1288	36.0	2760	6	AX765866 Sequence
28	1288	36.0	2760	8	BT002702 Arabidops
29	1288	36.0	3084	8	AY096797 Arabidops
30	1244	34.8	55839	1	AJ586576 Xanthomon
31	1228	34.4	26309	6	CQ363747 Sequence
32	1228	34.4	110000	1	AE017283 Continuation (22 o
33	1118	31.3	104045	2	AP006497 Continuation (9 of
34	1118	31.3	110000	2	AP006497 Continuation (8 of
35	1096.5	30.7	4496	6	A48326 Sequence 4
36	1096.5	30.7	4496	6	AR198356 Sequence
37	1068.5	29.9	349895	1	BX248359 Coryneb
38	1029	28.8	38829	2	AC151926 Phaeodact
39	937.5	26.2	298050	1	AP003189 Clostridi
40	930	25.3	11639	1	AE010477 Pseudocete
41	904.5	25.3	107097	8	AC005727 Arabidops
42	897	25.1	110000	8	CR382127 Continuation (6 of
43	859.5	23.1	10226	1	AE006363 Lactococ
44	837	24.1	110000	1	AE016822 Continuation (13 o
45	834	23.3	305050	1	AL596173 Listeria

## ALIGNMENTS

RESULT 1  
BD178313  
LOCUS  
DEFINITION Transformant producing pP1022 substance and process for producing the same and novel biosynthetic gene.  
ACCESSION BD178313.1 GI:30015578  
VERSION BD178313.1  
KEYWORDS WO 02077244-A/1.  
SOURCE Streptomyces venezuelae  
ORGANISM Streptomyces venezuelae  
REFERENCE Yanai, K., Sumida, N., Watanabe, M., Moriya, T. and Murakami, T.  
AUTHORS Streptomyces venezuelae  
TITLE Bacteria; Actinobacteria; Actinomycetales; Actinomycetales; Streptomyces; Streptomyces; Streptomyces; Streptomyces.  
JOURNAL 1 (bases 1 to 2061)  
PATENT Patent: WO 02077244-A 1 03-OCT-2002;  
MEIJI SEIKA KAISHA LTD, KOJI YANAI, NAOMI SUMIDA, MANABU WATANABE, TATSUKI MORIYA, TAKESHI MURAKAMI

REFERENCE Yanai, K., Sumida, N., Watanabe, M., Moriya, T. and Murakami, T.  
AUTHORS Streptomyces venezuelae  
TITLE Bacteria; Actinobacteria; Actinomycetales; Actinomycetales; Streptomyces; Streptomyces; Streptomyces; Streptomyces.  
JOURNAL 1 (bases 1 to 2061)  
PATENT Patent: WO 02077244-A 1 03-OCT-2002;  
MEIJI SEIKA KAISHA LTD, KOJI YANAI, NAOMI SUMIDA, MANABU WATANABE, TATSUKI MORIYA, TAKESHI MURAKAMI

COMMENT  
OS Streptomyces venezuelae  
PN WO 02077244-A/1  
PD 03-OCT-2002 WO 2002P002782  
PF 22-MAR-2002 WO 2002P002782  
PR 22-MAR-2001 JP 01P 082227  
PI KOJI YANAI, NAOMI SUMIDA, MANABU WATANABE, TATSUKI MORIYA, TAKESHI MURAKAMI  
PC C12N15/61, C12N1/15, C12P17/14, C12N1/15, C12R1/645, C12P17/14,



PC C12R1.645)  
CC Transformant producing PF1022 substance and process for CC  
CC producing the same  
CC and novel biosynthetic gene  
FH Key Location/Qualifiers  
FT CDS (1)..(2058).

FEATURES  
source  
1..2061  
/organism="Streptomyces venezuelae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:54571"

## ORIGIN

## Alignment Scores:

Pred. No.: 2..06e-166 Length: 2061  
Score: 3573.00 Matches: 686  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-10-089-514-2 (1-686) x BD178313 (1-2061)

QY 1 MetArgThrLeuLeuIleAspSerPheThrHisAsnLeuPheGlnTyrIle 20  
DB 1 ATGCGACGCTTCTGATCGACAACTACGACTCGTTACCCACAACTGTTCCAGTACATC 60  
QY 21 G1G1A1A1A1Thrg1Yg1InpProValValProAsnAspAlaAspTrpSerArgLeu 40  
DB 61 GGGAGAGCCACCGGGACACCCCGGTGCTGTCGCCACGACCGCATGCTGGTGGCGAGT 120  
QY 41 ProValG1uAspPheAspAlaIleValIseProG1yProG1ySerProAspArgGlu 60  
DB 121 CCGGTGAGAGACTTCAGCGGATGTCGTGTCCTCCCGGGCCCGGACGCCGACCGGAA 180  
QY 61 ArgAspPheG1yIleSerArgArgAlaIleThrAspSerG1yLeuProValIleuG1yVal 80  
DB 181 CGGAGACTTCGGAATCAGCCGCCGGGAGATCACGACAGCGGCTGCTCGCGCTC 240  
QY 81 CysLeuG1yHisG1ng1yIleAlaG1nLeuPheG1yG1yThrValG1yLeuAlaProG1n 100  
DB 241 TGGCTGGCCACCGAGGCATGCCCAAGCTTTCGGGGAAACGTCGAGCTTCGCCGAA 300  
QY 101 ProMetHisG1yArgValSerG1uValArgHisThrG1yG1uAspValPheArgG1yLeu 120  
DB 301 CCATGCAAGGCGGGGTCTCCAGGTGCGSCACACCGGAGGACGTCCTTCGGGGCTC 360  
QY 121 ProSerProPheThrAlaValArgTyrHisSerLeuAlaIleThrAspLeuProAspG1n 140  
DB 361 CCTCGCCGTTCAACCCCGCTGCTACACTCCCTGGCGCCACGACCTCCCGACGAG 420  
QY 141 LeuG1uProLeuAlaTrpSerAspAspG1yValMetG1yLeuArgHisArgG1yLys 160  
DB 421 CTGGAACCCCTCGCTGGAGCGACGAGGGGTGTCATAGGGCTTCGGGCAACCGGAGAG 480  
QY 161 ProLeuTrpG1yValG1nPheHisProG1uSerIleG1ySerAspPheG1yArgG1uIle 180  
DB 481 CCGCTGTGGGGCTTCCAGTTCCACCGGACTCCATCGGACGACACTTCGGCGGGAGATC 540  
QY 181 MetAlaAsnPheArgAspLeuAlaLeuAlaHisHisArgAlaArgHisG1yAlaAsp 200  
DB 541 ATGGCAAACTTCGCGACCTCGCCCTCGCCACCAACGGGACAGGCGCCACGGGGCCGAC 600  
QY 201 SerProTyrG1uLeuHisValArgArgValAspValLeuProAspAlaG1uG1uValArg 220  
DB 601 TCCCTGTACAACTTCACGTCGCGCGCTGACGTCTCGGACCCGAAAGGATACG 660  
QY 221 ArgG1yCysLeuProG1yG1uG1yThrPheTrpLeuAspSerSerValIleuG1n 240  
DB 661 CGGGCTGCTGCGCGGAGGAGGACACGTTCTGGCTGACAGCAGCTCCGTCCTCGAA 720  
QY 241 G1yAlaSerArgPheSerPheLeuG1yAspAspArgG1yProLeuAlaG1uTyrLeuThr 260

DB 721 GGGCGCTCGCGCTTCTCTCTCGGCGACGACCGGGCCCGCTCGCGAGTACCTCACCC 780  
QY 261 TyrArgValAlaAspG1yValValSerValArgG1ySerAspG1yThrThrArgTrp 280  
DB 781 TACCGGTGGCGAGCGGCTGCTCTCCGCGGCTCCGAGGACACCGACCGGAGG 840  
QY 281 ArgArgProPhePheAsnTyrLeuG1uG1uG1nLeuG1uArgArgValProValAla 300  
DB 841 CCGCGCCCTCTTCAACTACCTGGAGGAGCAGCTGAAACCGCGAGCGGCTCCCGTCCGC 900  
QY 301 ProG1uLeuProPheG1uPheAsnLeuG1yTyrValG1yTyrLeuG1yTyrG1uLeuLys 320  
DB 901 CCGGAATCGCTTCAGATTCAACTCGGCTGATGTCGCTACCTCGGCTACGAGCTGAAG 960  
QY 321 AlaG1uThrThrg1yAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe 340  
DB 961 GGGAGACACCGGCGACCCCGGACCGGTTCGCCGACCCGACCGCGCTTCTCTTC 1020  
QY 341 AlaAspArgAlaIleAlaLeuAspHisG1ng1yG1yCysTyrLeuLeuAlaLeuAsp 360  
DB 1021 GCGACCGGCGCATGCGCTCGACACAGAAAGGCTGCTGCTACCTGCGCTCGCCGAC 1080  
QY 361 ArgArgG1yHisAspAspG1yAlaArgAlaTrpLeuArgG1uThrAlaG1uThrLeuThr 380  
DB 1081 CGCGGGGCGACAGACGCGCGCGCTGCTGCGGAGACGCGCCGACGACCTTCACCC 1140  
QY 381 G1yLeuAlaValArgAlaProAlaG1uProThrProAlaMetValPheG1yIleProG1n 400  
DB 1141 GGGCTGGCTGCGGCGCGCGCGCGGCGAGCCGACCCCGCATGCTCTTGGAGATCCCCGAG 1200  
QY 401 AlaAlaAlaG1yPheG1yProLeuAlaArgAlaArgHisAspLysAspAlaTyrLeuLys 420  
DB 1201 GGGCGCGCGGCTTGGCGCCCTCGGCGCGGCGCGCGCACGACAGGACGCTTACCTCAAG 1260  
QY 421 ArgIleAspG1uCysLeuLysG1uIleArgAsnG1yG1uSerTyrG1uIleCysLeuThr 440  
DB 1261 CCGATCGAGAGATGCTTCAAGAGATCCCGCAACGGGAGTCTGTACGAGTTCGCTGACC 1320  
QY 441 AsnMetValThrAlaProThrG1uAlaThrAlaLeuProLeuTyrSerAlaLeuArgAla 460  
DB 1321 AACATGTCACCGCGCGACCGAGCGAGCGGCTTCGCTTATCTCCGCGCTCGCGCC 1380  
QY 461 IleSerProValProTyrG1yAlaLeuLeuG1uPheProG1uLeuSerValIleuSerAla 480  
DB 1381 ATCAGCCCGCTCCCGTACGCGCCCGCTCGATTCGCCGAATCGTGCCTGAGCGCC 1440  
QY 481 SerProG1uArgPheLeuThrT1leG1yAlaAspG1yG1yValG1uSerTyrProIleLys 500  
DB 1441 TCGCCCGAGCGGTTCTTCAAGATCGCGCGCGACGGCGGCTGAGTCCAAAGCCCATCAAG 1500  
QY 501 G1yThrArgProArgG1yG1yThrAlaG1uG1uAspG1uArgLeuArgAlaAspLeuAla 520  
DB 1501 GGGACCGCGCCCGGGGGGCGACCGGAGAGAGAGAGAGGCTCGCGCCGACCTGGGCC 1560  
QY 521 G1yArgG1uLysAspArgAlaG1uAsnLeuMetIleValAspLeuValArgAsnAspLeu 540  
DB 1561 GGGCGGGAAGAAGACCGGGCCGAGAACTGATGATGTCGACTGGTCCGCAACGACCTTC 1620  
QY 541 AsnSerValCysAlaIleG1ySerValHisValProArgLeuPheG1uValG1uThrTyr 560  
DB 1621 AACAGGTCGTGGCGCATCGGCTCCGTCACGTCGCCCGGCTCTTCGAGGTGAGACCTTAC 1680  
QY 561 AlaProValHisG1uLeuValSerThrIleArgG1yArgLeuArgProG1yThrSerThr 580  
DB 1681 GGGCCCGTACACAGTGTGTGTCACCATCCGGAGCGGCTCGGCGCCGACGACACC 1740  
QY 581 AlaAlaCysValArgAlaAlaPheProG1yG1ySerMetThrG1yAlaProLysLysArg 600  
DB 1741 GCGCGCTGGTACGCGCGCGCTTCCCGCGGCTCATACCGGCGCGCCAAAGAGCCG 1800  
QY 601 ThrMetG1uIleIleAspArgLeuG1uG1uG1yProArgG1yValTyrSerG1yAlaLeu 620  
DB 1801 ACCATGAGATCATGACCGCTGGAGGAAGGCCCGGGGCGTCTACTCCGGGCGCTC 1860



OY	621	GIJYTPHPbaLaLeuSerGIyAlaLaLeuSerLLevalIlaRgThrrLLevalLeu	640		
Db	1861	GGATGCTTCGCCCTCAGCGCGGCCGCCCACTCAGATCGATCGAACAATCGTGTG	1922		
OY	641	AlaAepGIyGlnAlaIGluPheGIyValGIyGlyAlaIleValSerLeuSerAaspGlnGlu	660		
Db	1921	GGCGACGGCCAGCGAGTTCGGCGTCGGCGGGGAGTCGTGTCTCTCCGACCGAGAG	1980		
OY	661	GIJGUluphethrGluThrValValValyAlaAgaIameValThrAlaLeuAaspGlySer	680		
Db	1961	GAGGAGGTTTCACCGAACAACCGTGTAAAGGCCGCCGACATGTCACCGCCCTCGACGCGAGC	2040		
OY	681	AlaValAlaIGlyAlaArg	686		
Db	2041	GCCGTGGCGGGCGGCCGA	2058		
RESULT 2	BD093914	2061 bp	DNA	linear	PAT 27-AUG-2002
LOCUS	BD093914				
DEFINITION	Transformants that produce secondary metabolites modified by a functional group(s) and novel biosynthesis genes.				
ACCESSION	BD093914	GI:22639502			
VERSION	BD093914.1	GI:22639502			
KEYWORDS	WO 0123542-A/1.				
SOURCE	Streptomyces venezuelae				
ORGANISM	Streptomyces venezuelae				
REFERENCE	Bacteria; Actinobacteridae; Actinobacteriales; Streptomycetales; Streptomycetaceae; Streptomycetes.				
AUTHORS	1 (bases 1 to 2061) Yanai,K., Okakura,K., Yasuda,S., Watanabe,M., Miyamoto,K., Mido,N. and Murakami,T.				
TITLE	Transformants that produce secondary metabolites modified by a functional group(s) and novel biosynthesis genes				
JOURNAL	Patent: WO 0123542-A 1 05-APR-2001; MEIJI SEIYA KAISEI LTD, KOJI YANAI, KAORU OKAKURA, SHOHEI YASUDA, MANABU WATANABE, KOICHI MIYAMOTO, NAOKI MIDO, TAKESHI MURAKAMI				
COMMENT	OS Streptomyces venezuelae PN WO 0123542-A/1 PD 05-APR-2001 PF 29-SEP-2000 WO 2000JP006783 PR 29-SEP-1999 JP 99P 276314 PI KOJI YANAI, KAORU OKAKURA, SHOHEI YASUDA, MANABU WATANABE, KOICHI MIYAMOTO, PI NAOKI MIDO, TAKESHI MURAKAMI PC C12N15/09, C12N5/10, C12P21/02, C07K11/00/(C12P21/02, C12R1:645) CC FH FT CDS				
FEATURES	source				
ORIGIN	1..2061 /organism="Streptomyces venezuelae" /mol_type="genomic DNA" /db_xref="taxon:54571"				
Alignment Scores:	2,066-166	Length:	2061		
Pred. No.:	3573.00	Matches:	686		
Score:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatch:	0		
Query Match:	100.00%	Indel:	0		
DB:	6	Gaps:	0		
US-10-089-514-2 (1-686) x	BD093914 (1-2061)				
OY	1	MeArghThrLeuLeuIleAaspAsnTyRaspSerPheThrhiAsnLeuPheGlnTyIle	20		
Db	1	ATGGCGACGCTTCATGCAGCAACTACGACATCGTTACCCCAACCTGTTCCAGACATC	60		
OY	21	GIJGUlualAThngIyGlnProProValValValProAsnAaspIlaAspTrpSerArgLeu	40		
Db	61	GAGGAGGCGACCGGGCAACCCCGCTGTGTGTGGCCCAACGCGGACTGTGTGGCGGTG	120		

QY	41	ProValGIuaSPheAspA1a1IleValSerProGIYserProAspArgIu	60
Db	121	CCCGTCGAGGACTTCGAGCGAGTGTCTGTCCCGGGCCCGGAGGCCGACCGGGAA	180
QY	61	ArgAspPheGIYIleSerArgA1a1IleThrAspSerGIYLeuProValLeuGIYVal	80
Db	181	CGGAGCTTCGGAAATCAGCCGCCCGGGGCATCACCGAGCGGCTCCGCTCCGGCGT	240
QY	81	CysLeuGIYHisGIYIleAlaG1mLeuPheGIYGIYThrValGIYLeuA1aProGIu	100
Db	241	TGCCTCGGCGCACAGGCGCATGCCACGCTCTTCGGCGGAACGTGTGGCTCGCCCCGGAA	300
QY	101	ProMetHisGIYArgValSerGIYValArgHisThrGIYThrGIYValSerValPheArgGIYLeu	120
Db	301	CCCATGACAGGCGCGAGTCTCCGAGGTGGGACACCGCGAGAGAGTCTTCGGGGGCTTC	360
QY	121	ProSerProPheThrAlaValArgTYHisSerLeuA1a1IleThrAspLeuProAspGIu	140
Db	361	CCCTCGCGGTTCACCGCGGTGGCTACACACTCTGGCGGCGCACCGACTCCCGACGAG	420
QY	141	LeuGIuProLeuA1aTPSerAspAspGIYValValMetGIYLeuArgHisArgGIuLys	160
Db	421	CTCGAACCCCTCGCTGGAGCGAGACGCGGTCTGTATGGCTTCGGCGCACCGCGACGAG	480
QY	161	ProLeuTPRG1YValG1nPheHisProGIuSerIleGIYSerAspPheGIYArgGIuIle	180
Db	481	CCGCTGTGGGCGTCCAGTTCACCGGAGTCCATCGGACAGCATTCGGCGGGAGATC	540
QY	181	MetAlaAspPheArgAspLeuA1aLeuA1aHisHisArgA1aArgArgHisGIYAlaAsp	200
Db	541	ATGGCCAACTTCGCGACCTCGCTCGCCGCCACACACCGGCGCACGGGGCCGAC	600
QY	201	SerProTYGIuLeuHisValArgArgValAspValLeuProAspA1aGIuGIuValArg	220
Db	601	TCCCCGTAAGCAATCCACGTGCGCGGTCTCACCTGTGCCGAGCGCGGAAGAGGTACGC	660
QY	221	ArgGIYCysLeuProGIYGIuGIYThrThrPheThrLeuAspSerSerValLeuGIu	240
Db	661	CGCGGCTGCTCCCGGAGGAGGACACAGTTCGGCTGGACAGAGCTCCGTCTCGAA	720
QY	241	G1YAlaSerArgPheSerPheLeuGIYAspAspArgGIYProLeuA1aGIYTYrLeuThr	260
Db	721	GGGCGCTCGGCTTCTCTCTCTCGGGAGAGACCGGGCCCGCTCGCGAATCACTCAC	780
QY	261	TYrArgValAlaAspGIYValValSerValArgGIYSerAspGIYThrThrArgThr	280
Db	781	TACGCGCTCGCGGACGCGTCTCGCTCGCGGCTTCGACGCGCACACGACGCGGAGC	840
QY	281	ArgArgProPhePheAsnTYrLeuGIuGIuGIuLeuGIuArgArgArgValProValAla	300
Db	841	CGGCGCCCTTCTTCACTACTCGAGAGACGACTCGAACGCCAGCGGATCCGCTCC	900
QY	301	ProGIuLeuProPheGIuPheAsnLeuGIYTYrValGIYTYrLeuGIYTYrGIuLeuLys	320
Db	901	CCGGAACCTGCGCTTCGAGTTCAACTCGGCTACGTCCGCTACCTCGGCTACGAGCTGAG	960
QY	321	AlaGIuThrThrGIYAspProAlaHisArgSerProHisProAspA1aIaPheLeuPhe	340
Db	961	GCGGAGACCAACCGGCGACCCCGGCGACCGGTCCCGGACCCCGACGCGCTTCTTC	1020
QY	341	AlaAspArgA1aIleAlaLeuAspHisGIuGIuGIYCysCysTYrLeuLeuA1aLeuAsp	360
Db	1021	GCGGACCGCGCATCGCTTCGACACACGAGAGAGCTGTCTACTGTCTGTGCTCGAC	1080
QY	361	ArgArgGIYHisAspAspGIYAlaArgA1aTPLeuArgGIuThrAlaGIuThrLeuThr	380
Db	1081	CGCGGGGCGACAGCGAGCGCGCGCTGTGGCTGGGAGAGCGCGGACCTCAAC	1140
QY	381	GIYLeuA1aValArgA1aProAlaGIuProThrProAlaMetValPheGIYIleProGIu	400
Db	1141	GCGCTGGCGGTCCGCGCCCGGCGGACGACCCCGCATGTGTCTTCGGATCCCGGAG	1200







PMNQGRVSLVKDRAARYAADHGLDSEFLVNLVDVITTEMCRVEDLVMSSESLTAEDR  
R"  
4065..5243  
/gene="ORF1V"  
4065..5243  
/gene="ORF1V"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAD21143.1"  
/db\_xref="GI:47846869"  
/translation="MATTTTAAALTGRTAAQPPYMPDPAVDIVTERELAGLPVLTPE  
EVDLMNGMALVVRGAALLVGGGCAARFPEAVPDIVRRVNDLQGLAATMRAGSGLP  
AVALRAGQVGRKSPFETESGAKRKPSPCGAVNRPPEPARPTEPRRLTLA  
YTCQSQIVLDEIRSRMSGRPLYERYTTSHELLLPYEPFLVREGAHGITSASAFMGWIG  
ERTREDEAHVALAQAVHNPVGVKGLPTVSPEDAVALSRINLEGVAGVRLTFLVRCA  
KEVDELLPVRVAVARHGAIVMLCDPMHNGDKLAGHKTRLIEPMRAETAAEVRTIR  
EHQWMPAGLHELTPDPTVECELDPRPFRFTDYRSTCDRLNPEQADWVTHFLSLT  
"

## ORIGIN

Alignment Scores:  
Pred. No.: 5,72e-166 Length: 5251  
Score: 3573.00 Matches: 686  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 1 Gaps: 0

US-10-089-514-2 (1-686) x AB116234 (1-5251)

QY 1 MetArgThrIleuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGlnTyrIle 20  
DB 651 ATCCGACCGCTTCTGATCGAACACTAGACTCGTTCCACCAACCTGTTCCAGTACATC 710  
QY 21 G1G1uAlaThrG1yGlnPProValIValProAsnAspAlaAspTTPSerArgLeu 40  
DB 711 GGGAGGCGCACCGGCGAACCCCGCTGCTGCTGCCACACGCGCACTGTCGGCGCTG 770  
QY 41 ProValG1uAspPheAspAlaIleValIleSerProG1yProG1ySerProAspArgLeu 60  
DB 771 CCCGTCGAGGACTTCCACGCGATCGTCTGCCCGGCGCGGCGAGCCCGACCGGAA 830  
QY 61 ArgAspPheG1yIleSerArgArgAlaIleThrAspSerG1yLeuProValIleG1yVal 80  
DB 831 CGGAGCTTCGGAATCAGCCCGCGGCGATCACCGACAGCGGCTGCTCGGCGCTC 890  
QY 81 CyeLeuG1yHisGlnG1yIleAlaGlnLeuPheG1yG1yThrValG1yLeuAlaProG1y 100  
DB 891 TGCCTCGGCGCACAGGCGATCGCCAGCTTTCGGCGGAACCGTCCGCTCGCCCGGAA 950  
QY 101 ProMetHisG1yArgValIleSerG1yValIleArgHisThrG1yG1uAspValPheArgG1yLeu 120  
DB 951 CCCATCGACGCGCGGCTCTCCAGGTGCGGACACCGGCGAGGACGCTTTCGGGCGCTC 1010  
QY 121 ProSerProPheThrAlaValIleArgTyrHisSerLeuAlaIleThrAspLeuProAspG1y 140  
DB 1011 CCTTCGCCCTTCCACCGCCGCTGCTCACTCCCTGCGCCGACCACTCCCGCGAG 1070  
QY 141 LeuG1uProLeuAlaIleTrpSerAspAspG1yValIleMetG1yLeuArgHisAspG1yVal 160  
DB 1071 CTCGAACCCCTCGCTCGAGCGACGAGCGGGGTGTCATGGGCTCGGCGACCGCGAGAA 1130  
QY 161 ProLeuTTPG1yValGlnPheHisProG1ySerIleG1ySerAspPheG1yArgG1yIle 180  
DB 1131 CGGCTGCGGCGCTCCAGTTCCACCGGAGTTCATCGGACGAGCACTTTCGGCGGAGATC 1190  
QY 181 MetAlaAsnPheArgAspLeuAlaLeuAlaHisHisArgAlaArgArgHisG1yValAsp 200  
DB 1191 ATGGCCAACTTCGCGACCTCGCTCGCCACACCGGCGACGCGGCGCGACCGGCGCGAC 1250  
QY 201 SerProTyrG1uLeuHisValIleArgValAspValIleuProAspAlaG1uG1yValArg 220

DB 1251 TCCCGGTACGAACCTCCACGTCGCCCGCGCTCGAGTCTGCCGACCGCGAAGAGTACGC 1310  
QY 221 ArgG1yCysLeuProG1yG1uG1yThrPheTrpLeuAspSerSerValIleG1y 240  
DB 1311 CGCGGCTGCTGCGCCGCGAGGAGCACCACTTCTGCTGGACAGACGACTCGTCTCGAA 1370  
QY 241 G1yAlaSerArgPheSerPheLeuG1yAspAspArgG1yProLeuAlaG1yTyrIleuThr 260  
DB 1371 GGGCTCGCGCTTCTCTCTCTCGCGAGACGACCGGCGCGCTCGCGTACCTTCACC 1430  
QY 261 TyrArgValAlaAspG1yValIleSerValIleArgIleSerAspG1yThrThrArgTyr 280  
DB 1431 TACCGGTCCCGCACGCGCTCGTCTCGCTCGCGGCTCCGACCGCACACGACCGGACG 1490  
QY 281 ArgArgProPhePheAsnTyrIleuG1uG1uG1uArgArgValProValAla 300  
DB 1491 CGCGCGCCCTTCTTCAACTCACTCGAGGAGACGCTGAAACGCGGCTCGCGTCCGCTGCC 1550  
QY 301 ProG1uLeuProPheG1uPheAsnLeuG1yTyrValG1yTyrIleuG1yTyrG1uLeuVal 320  
DB 1551 CCCGAACCTCGCTTTCAGATTCAACTCGGCTAGCTGCTGCTTCACTCGGCTGAG 1610  
QY 321 AlaG1uThrThrG1yAspProAlaHisAspSerProHisProAspAlaAlaPheLeuPhe 340  
DB 1611 GCGGAGACCAACCGGCGACCCCGCGCACCGGTCCCGCACCCGACCGCGCTTCTCTTC 1670  
QY 341 AlaAspArgAlaIleAlaLeuAspHisGlnG1yG1yCysArgTyrIleuLeuAlaLeuAsp 360  
DB 1671 GCCGACCGGCGCATCGCTCGACCAACGAGAAAGCTGCTGCTCACTGCTGCGCTCGAC 1730  
QY 361 ArgArgG1yHisAspAspG1yAlaArgAlaIleTrpLeuArgG1uThrAlaG1uThrLeuThr 380  
DB 1731 CGCGGAGGCGACGACACGCGCGCGCTGCTCGGAGACGCGCGACCTTACCC 1790  
QY 381 G1yLeuAlaValArgAlaProAlaG1uProThrProAlaMetValPheG1yIleProG1y 400  
DB 1791 GGCCTGCGCTCGCGCGCGCGCGCGAGCGACCCCGCGCATGCTTTCGGATCCCGAG 1850  
QY 401 AlaAlaAlaG1yPheG1yProLeuAlaArgAlaIleArgHisAspIleAspAlaIleTyrLeuVal 420  
DB 1851 GCGCGCGCGCTTGGCGCCCGCTGCGCGCGCGCGCGCACGACGACCGCTTACCTCAAG 1910  
QY 421 ArgG1yAspG1uCysLeuLeuG1yIleArgAsnG1yG1uSerTyrG1uIleCysLeuThr 440  
DB 1911 CGCATCGAGAGGCTCTCAAGAGATCCGCAACGCGGAGTCTGACAGATCTGCTGACC 1970  
QY 441 AsnMetValThrAlaProThrG1uAlaThrAlaLeuProLeuTyrSerAlaLeuArgAla 460  
DB 1971 AACATGTCACCGCGCGACCGAGCGAGCGCCCTGCGCTTCACTCCGCGTGGCGCC 2030  
QY 461 IleSerProValProTyrG1yAlaLeuLeuG1uPheProG1yLeuSerValIleuSerAla 480  
DB 2031 ATCAGCCCGCTCCGTAACGCGCCCTGCTCGAGTTCGCCGAATGTCGTCGAGCGCC 2090  
QY 481 SerProG1uArgPheLeuThrIleG1yAlaAspG1yG1yG1uSerIlePheIleVal 500  
DB 2091 TCGCCCGAGCGGTTCTCTCAAGATCGGCGCGAGCGGCGGTGAGTCCAAAGCCATCAAG 2150  
QY 501 G1yThrArgProArgG1yG1yThrAlaG1uG1uAspG1uArgLeuArgAlaAspLeuAla 520  
DB 2151 GGGACCGCGCCCGCGGCGCGCACCGCGAGAGAGACGAGCGGCTCGCGCCGACCTGGCC 2210  
QY 521 G1yArgG1uIleAspArgAlaG1uAsnLeuMetIleValAspLeuValArgAsnAspLeu 540  
DB 2211 GCGCGGAGAAAGAACCGGCGAGAACCTGATGATGTCGATCTGTCGCGACGACCTTC 2270  
QY 541 AsnSerValCysAlaIleG1ySerValHisValProArgLeuPheG1uValG1uThrTyr 560  
DB 2271 AACAGGTCTGCGCATCGGCTCGCTCCACGTCGCCCGCTTTCGAGGTGAGACCTTAC 2330  
QY 561 AlaProValHisG1uLeuValSerThrIleArgG1yArgLeuArgProG1yThrSerThr 580



Db 2331 GCCCCGTCACCACTGGTGTGACCACTCCGGGAGCGCTGGCCCGGCACCAAGCACC 2390  
Qy 581 ALaAlaCySeValaArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLysLysArg 600  
Db 2391 GCCGCTGCGTACGCGCCGCTTCCCCGGGGGCTCCATGACCGCGCGCCCAAGAACGC 2450  
Qy 601 ThMetGluLeileleapArgLeuGluGlyProArgGlyValTyrSerGlyAlaLeu 620  
Db 2451 ACCATGAGATCATCAACCCCTCGAGAGAGCGCCCGGGGCGTCACTCCGGGCGCTC 2510  
Qy 621 GlyTyrPheAlaLeuSerGlyValAlaAlaSerLeuSerLeileValTyrTrilleValLeu 640  
Db 2511 GGATGGTTCGCTCGAGCGCGCGCGCTGACCTGATCGTCACTCCGACCATCGTCTG 2570  
Qy 641 AlaAspGlyGlnAlaGluPheGlyValGlyGlyAlaileValSerLeuSerAspGlnGly 660  
Db 2571 GCCGACGCGCACGCGGAGTTCCGCGCTCGGGGGGCGATCGTCTCCTCCGACCAAGAG 2630  
Qy 661 GluGluPheThrGluThrValValAlaLysAlaArgAlaMetValThrAlaLeuAspGlySer 680  
Db 2631 GAGGAGTTCACGAGACCGGTAAAGCGCGCGCCATGTTCAACCGCCTCGACGGCAGC 2690  
Qy 681 AlaValAlaGlyValaArg 686  
Db 2691 GCCGTGCGGGCGCCCGA 2708

RESULT 4  
SVU21728 2660 bp DNA linear BCT 22-NOV-1996  
LOCUS Streptomyces venezuelae p-aminobenzoic acid synthase (pabAB) gene,  
DEFINITION complete cds.  
ACCESSION U21728  
VERSION U21728.1 GI:1680048  
KEYWORDS  
SOURCE Streptomyces venezuelae  
ORGANISM Streptomyces venezuelae  
Bacteria; Actinobacteridae; Actinomycetales;  
Streptomycinae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 2660)  
AUTHORS Brown,M.P., Aidoo,K.A. and Vining,L.C.  
TITLE A role for pabAB, a p-aminobenzoate synthase gene of Streptomyces  
JOURNAL venezuelae ISB5230, in chloramphenicol biosynthesis  
MEDLINE Microbiology 142 (Pt 6), 1345-1355 (1996)  
PUBMED 96262706  
REFERENCE 2 (bases 1 to 2660)  
AUTHORS Brown,M.P.  
TITLE Direct Submission  
JOURNAL Submitted (26-FEB-1995) Biology, Dalhousie University, Halifax,  
COMMENT Novae Scotia B3H 4H6, Canada  
FEATURES  
location/Qualifiers  
source 1..2660  
/organism="Streptomyces venezuelae"  
/mol\_type="genomic DNA"  
/strain="ISP5230"  
/db\_xref="taxon:54571"  
<1..507  
/note="ORF1"  
/codon\_start=1  
/transl\_table=1  
/product="unknown"  
/protein\_id="AAB30312.1"  
/db\_xref="GI:1680049"  
/translation="GSAEYVALPFLGHEGDLIRSKTGMIRRGSESVLPADSTNIS  
PMYITVAEYIGVDIVFVGESIGAAASWVGPLGEPIDRTDTSRLNGSNFPOA  
REIVDALEPDEYVYVAMGDEPMWGVMAVDYESHDAIVDSDLVAHVODKGTARL  
HURRLRL"  
gene 651..2660  
CDS /gene="pabAB"  
651..2660  
/gene="pabAB"  
/codon\_start=1

/transl\_table=1  
/product="p-aminobenzoic acid synthase"  
/protein\_id="AAB30312.1"  
/db\_xref="GI:710438"  
/translation="NRITLIDNYDSFTONTLFOYIGENATQPPVVENDAWMSRLPLEDF  
DAIVSPGSGSPDRERDFEISRRALTDSGLPLGVCLEHQIAGLSAEPHMRSEVR  
HTGEIVFRGLPSPFTAVRYSLSAATDLDPLELEPLMSDGVMGRLHREKPLMGVQFP  
PESIGDFGRIEMANFRDLALHHRARADADMGVLEHYVRVDVLPDAEVRRAACPA  
EGATPELSSSVLEGASPPSFLGDDRGPLAEYLYRAVDGVSVRSDGTTTRDAATL  
FSLBEOLLEPPAGVPVAPDLPEFNLGVYGLGVELEKATETGDPAYAPHPDAAPLPAD  
RAIALDHQGCCYLLADBRGHDDARAMLRTATLTGLAARVRPRPRMVRCPVE  
AAAGPPLARARHDQASALRNSESITCLTMTVAPEATLPLYSALRRISVPVSG  
ALLPEPELSVLSASPERPLTIGADGVSKPLKGRPRGAPAEDEERLARADLAGKED  
RAENLMIVLVNDLNSVCAIGSVHVPRLFEVGDLAPYHQLVSTIRGLRPGTSTAC  
VRAAPFGSGMTCAPKRRPWEIIDRLSEEGRGVLPALGMLGASTGARRISVIRITVLA  
DGRARFGVGAIVLSLDSEERFRQIVVARAMVATLDGSAVAGAR"

ORIGIN

Alignment Scores:

	Pred. No.:	1..8e-144	Length:	2660
Qy	Score:	3127.50	Matches:	618
Db	Percent Similarity:	91.40%	Conservative:	9
	Best Local Similarity:	90.09%	Mismatches:	42
	Query Match:	87.53%	Indels:	17
DB:		1	Gaps:	3

US-10-089-514-2 (1-686) x SVU21728 (1-2660)

Qy 1 MetAgtThLeuLeuileaspAntyrraBserPheThHiasnLeuPheGlnTyrile 20  
Db 651 ATGCGACGCTTCTGTATGCAACAACGACTGTTTCAACCAAGCCTGTTCCAGATACATC 710

Qy 21 GlyGluAlaThGlyGlnProProValValAlaProasnAspAlaAspTyrSerArgLeu 40  
Db 711 GCGAGGCCACCGGGGAGCCCCCC--GTCGTGCCACAGACGCCGACTGTGCGGGCTG 767

Qy 41 ProValGluAspPheAspAlaileValaLysSerProGlyProGlySerProAspArgGlu 60  
Db 768 CCCCTGAGAGACTTCGACGCGATCGTGTCCCGGGGCCCGGGACGCCCGACCGGANA 827

Qy 61 ArgAspPheGlyIleSerArgArgAlaileThraAspSerGlyLeuProValLeuGlyVal 80  
Db 828 CCGGACTTGGGATACGCGCGGGGAGACCGACAGCGGCGCTGCCCTCCGCGCTC 887

Qy 81 CylLeuGlyHISGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaProGlu 100  
Db 888 TGCTTCGCGCCACAGGCGATCGCCAG-----CTTCGCGCGANA 926

Qy 101 PromethIleGlyArgValSerGluValaArgHistrGlyGluAspValPheArgGlyLeu 120  
Db 927 CCATGCAAGCGCGGGTCTCCGAGGTGGCGGACACCGGGGAGAGAGCGTTCCGGGGCTTC 986

Qy 121 ProSerProPheThrAlaValaArgTyrHisSerLeuAlaAlaThrAspLeuProAspGlu 140  
Db 987 CCTCGCCCTTACACGCGCGGTGCTACCACTCCGCGCGCACCGACCTCCCGACGAG 1046

Qy 141 LeuGluProLeuAlaTyrSerAspAspGlyValaMetGlyLeuArgHistrArgGlyLys 160  
Db 1047 CTGGAACCCCTCGCTCGAGCGACAGCGCGCTGTCATAGGCTCGGACCGCGAGAG 1106

Qy 161 ProLeuTyrGlyValaGlnPheHisProGluSerIleGlySerAspPheGlyArgGluile 180  
Db 1107 CCGCTGATGGGCGTCCAGTTCCACCGGAGTCCATCCGAGGAGACTTCGGCGGGAGATC 1166

Qy 181 MetAlaasnPheArgAspLeuAlaLeuAlaHistrHISArgAlaArgArgHistrGlyAlaAsp 200  
Db 1167 ATGGCCACTTCGCGCACTTCCTCGCGCACACCGGGGACGTCGCGACACGCGCGCAC 1226

Qy 201 SerProTyrGluLeuHistrValaArgArgValaAspValLeuProAspAlaGluGluValaArg 220  
Db 1227 TGGGCTTACGAACCTCCACGTGGCGCGGTGCGACGTGCTGCGGACCGCAAGAGGTACCC 1286

Qy 221 ArgGlyCyaleuProGlyGluGlyThrThrPheTyrLeuAspSerSerValLeuGlu 240



[illegible]

Db	2340	GCCGCTGCGTACGGCGCCGCTTCCCGGCGGCTTCATGACGGCGCGCCCAAGAAAGCA	239
Qy	601	ThnrtwgluilelleaspargluengluglylproargglyValTyrSerGlyAlaLeu	620
Db	2400	CCCATGAGATCATCGACCGCTCGAAGAAAGCCCCCGGCGGTCTTACCCGGGCGCTC	245
Qy	621	GlyTTPheAlaLeuSerGlyAlaAlaAspIeuSerIleValIleArgThrIleValLeu	640
Db	2460	GGATGGTTCGCCCTCACGGCGGCGCGGACCTCAGCATCGTCATCCGACCATCTGCTG	251
Qy	641	AlaAspIgluAlaIleaglupheGlyValGlyValIleAlaIleValSerLeuSerAspIngu	660
Db	2520	GCCACGGCGCGGCGCGAGTTTCGGCTCGGCGCGGCGATCGTCTCTTCCGACCGAGAG	257
Qy	661	GlulupheThrgIuThrnValValIleAlaArgAlaMetValThrnAlaLeuAspGlySer	680
Db	2580	GAGAGATTACGACGACGCGTGTCAAGGCGCGCCCATGTGTCACCGCTTCGACGGCAGC	263
Qy	681	AlaValAlaaglAlaIleArg	686
Db	2640	GCAGTGGCGGGCGCCCGA	2657
RESULT 5			
LOCUS	AF262220	14159 bp	DNA linear BCT 09-JAN-2002
DEFINITION	Streptomyces venezuelae chloramphenicol biosynthetic gene cluster, partial sequence.		
ACCESSION	AF262220		
VERSION	AF262220.2	GI:14290414	
KEYWORDS			
SOURCE	Streptomyces venezuelae		
ORGANISM	Streptomyces venezuelae		
REFERENCE	He,J., Magarvey,N.A. and Vining,L.C. The gene cluster for chloramphenicol biosynthesis in Streptomyces venezuelae ISP5230 includes novel shikimate pathway homologues and a monomolecular non-ribosomal peptide synthetase gene Microbiology 147 (Pt 10), 2817-2829 (2001)		
AUTHORS	He,J., Magarvey,N.A. and Vining,L.C.		
TITLE	Submitted (01-MAY-2000) Biology, Dalhousie University, 1355 Oxford St., Halifax, NS B3H 4J1, Canada		
JOURNAL	Submitted (05-JUN-2001) Biology, Dalhousie University, 1355 Oxford St., Halifax, NS B3H 4J1, Canada		
REFERENCE	Sequence update by submitter		
AUTHORS	On Jun 5, 2001 this sequence version replaced gi:10716943.		
TITLE	Location/Qualifiers		
JOURNAL	1. 14159		
REMARK	/organism="Streptomyces venezuelae"		
COMMENT	/mol_type="genomic DNA"		
FEATURES	/strain="ISP5230"		
SOURCE	/db_xref="taxon:54571"		
	/map="adjacent to pabAB"		
	2333..3511		
	/function="essential for chloramphenicol biosynthesis"		
	/codon_start=1		
	/transl_table=11		
	/product="CmIG"		
	/protein_id="AAG21974.1"		
	/db_xref="GI:10716944"		
	/translation="MTAGSPGPAGRTPPLRPALAVRDHSFLADIGMDQVNPCTAFV		
	GTEPEPVEDPAAWRREFVARGICGTALADLVAAEDVYVTLDPKPGVCGPAELAD		
	LVTAYNSRRLRIDEVRLAYRHHHTGEBYVAFKFSHLIYDLADVDDLQIRHLNG		
	DTPPEIGSYRHHARTVEQYVALPADPAEVYRALGELPVPGRKGIPTISASVEEVL		
	PLAEQTEFELLASVTGVLLPALGGGVIVLOYPSRWEFAFKGGYVEIKPLVVRPGILP		



```
CDS
3498..6604
/note="non-ribosomal peptide synthetase; contains
reductase-like domains near C-terminus"
/codon_start=1
/transl_table=11
/product="CmIH"
/protein_id="FAG21975.2"
/db_xref="GI:14327897"
/translation="MELPDLITDLPFRHGRRTLRAGRTWYEEIDRYTSALRIID
AECPPAGRRVIVAGEHTAEVVAALAAARSHAVHPNPGLPADRFEEPAVALALV
CFERBALVRGEKAGLRALYAGDVGPDPADPADGADBPAPASRGVAFISGSGD
PKLVDSHGGLNLCSRLRLDITDDQDLPAPSLPDSVSEILGTVAGATLVVP
VRDQASWLGVSVSHLAHGCRLAMLSVSVARLDFAARSRIRVEECGELGKGEYDK
AARYSRVNAVGPTEATVCGSLAMELSTYPTSTPYDPGPAIVRDPDSDHATGCE
LVIIVGQVALGVAGSSPAENEVFTVDGSAVATGVNLSIDGELTVIGRIDEQIKR
LGRHVNLAHGSTLSRHLGREVALVRQDAILLVTADGATEBSLMARTRDVPMWE
APRLVIVDALPTSGKVDRSALRELLSPAGAPHGDTGDEDAELRLVDVTVVL
GOELGPESTSIFDAGSSIAMIOIOVKLSDVGEAVEAFAAMDYOPAPAFRLHNG
EAVAPASAVDTLLRRVETERRDALRAELPLRRDTHHEPVGAAADGRVTLTGASGF
IGGHVLDRLAARPVILVSTGDDPVLTGHATRFGRQADYARVRAISYAEIERVD
RRGPPVDVAVHCGYQVNHLLPLDSHLSGSVNRNALVVRALAAAGARSAFSLERVD
ADRLPLSAATLAVGDVPSRKLISEYVNTLAVLGCAVSHRPGVLVGRPRDRHL
KDDMFALLETARVGMPLSGHPVPCDVCTGLADTLGRDPANPGTADGASRTPSR
SAVVHRTVALDELHRTGLTEADVLAAPAMFERVDRGEPVPLLAAMQALSGGW
PSAAREVDHDLGRLLGTPPTPDAGDRPRTGTSRGPERRRRADTALTGTPR"
misc_feature
3576..3596
/note="Region: adenylation domain A1"
misc_feature
3708..3746
/note="Region: adenylation domain A2"
misc_feature
3939..3986
/note="Region: adenylation domain A3"
misc_feature
4089..4100
/note="Region: adenylation domain A4"
misc_feature
4350..4370
/note="Region: adenylation domain A5"
misc_feature
4494..4538
/note="Region: adenylation domain A6"
misc_feature
4593..4610
/note="Region: adenylation domain A7"
misc_feature
4647..4706
/note="Region: adenylation domain A8"
misc_feature
4833..4853
/note="Region: adenylation domain A9"
misc_feature
4893..4910
/note="Region: adenylation domain A10"
misc_feature
5043..5075
/note="Region: chiotation domain"
misc_feature
5343..5375
/note="Region: NADH binding site"
ORIGIN
Alignment Scores:
Pred. No.: 1,12e-143 Length: 14159
Score: 3127.50 Matches: 618
Percent Similarity: 91.40% Conservative: 9
Best Local Similarity: 90.09% Mismatches: 42
Query Match: 87.53% Indels: 17
DB: 1 Gaps: 3
US-10-089-514-2 (1-686) x AF262220 (1-14159)
QY 1 MetArgThrIleuLeuIleAspAsnTYrAspSerPheThrHisAsnLeuPheGlnTYrIle 20
DB 8153 ATGGGACGCGTTCTGATCGACAACACTACGACTCGTTCCACCAAGAACTGTTCCAGTACATC 8212
QY 21 GlyGluAlaThrGlyGlnProProValValAlProAsnAspAlaAspTTPSerArgLeu 40
DB 8213 GGGGAGGCGCAACGGGCGCCCCCCC--GTGCTGGCCCAACGACGCGGCTGCTGCTGCGGCTG 8269
QY 41 ProValGluAspPheAspAlaIleValValSerProGlyProGlySerProAspArgGlu 60
```

```
DB 8270 CCCCTCGAGAGACTTCACACCGATCGCTGTCGCCCGGCGCCGCGGACCCGACCGCGAA 8329
QY 61 ArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeuGlyVal 80
DB 8330 CGGACCTTCGGANTCAGCCGCGGCGATCATCCGACAGCGGCGCTCCCTCCGCGCTC 8389
QY 81 CysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaProGlu 100
DB 8390 TGCCTCGGCGACACGAGGATCCGCCAG-----CTCTCGGCGGAA 8428
QY 101 PromethiGlyArgValSerGlyValAlaArgHisThrGlyGluAspValPheArgGlyLeu 120
DB 8429 CCCATCAGCGCGCGGCTCTCCAGAGCGCGCACCGCGGAGAGACTTCCGCGGCGCTC 8488
QY 121 ProSerProPheThrAlaValArgTYrHisSerLeuAlaAlaThrAspLeuProAspGlu 140
DB 8489 CCTTCGCGCTCACCCGCTGCGCTACACTCCCTCGGCGGACGACACTCCCGACGAG 8548
QY 141 LeuGluProLeuAlaTrpSerAspAspGlyValAlaMetGlyLeuArgHisArgGluLys 160
DB 8549 CTGGAACCCCTCGCTCGAGCGACGACGCGCTGCTCATGGGCTCGCGGACCGCGAGAG 8608
QY 161 ProLeuTrpGlyValGlnPheHisProGlySerIleGlySerAspPheGlyArgGluIle 180
DB 8609 CCGGTATGGGCGTCCAGTTCCACCGGAGTCCATCGGACGACGACTTCCGCGCGGAGATC 8668
QY 181 MetAlaAsnPheArgAspLeuAlaLeuAlaHisHisAlaArgAlaArgArgHisGlyAlaAsp 200
DB 8669 ATGGCCAACTTCGCGCACTCCGCTCCGCGCACACCGGCGACGCTGCGACCGCGCGAC 8728
QY 201 SerProTYrGlyLeuHisValArgArgValAspValLeuProAspAlaGluGluValArg 220
DB 8729 TGGGCGTACGAACCTCCACCTGCGCGCTCGACGCTCCTCCGACCGCGCAAGAGTACC 8788
QY 221 ArgGlyCysLeuProGlyGlyGlyThrThrPheTrpLeuAspSerSerValLeuGlu 240
DB 8789 CGGCGTCCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8848
QY 241 GlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTYrLeuThr 260
DB 8849 GGGCGCTCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8908
QY 261 TYrArgValAlaAspGlyValValSerValArgGlySerAspGlyThrThrArgTYr 280
DB 8909 TACCGCGTCCGCGGCGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8968
QY 281 ArgArgProPhePheAsnTYrLeuGluGluGluGluArgArgValProValAla 300
DB 8969 GCGGCGACCTCTTCACTACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9028
QY 301 ProGluLeuProPheGluPheAsnLeuGlyTYrValGlyTYrLeuGlyTYrGluLeuLys 320
DB 9029 CCCGACCTGCGCTTCAAGTTCAACTCGGCTCTGCTGCTCTCTCTCTCTCTCTCTCTCT 9088
QY 321 AlaGluThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe 340
DB 9089 GCGGAGACCAACCGGCGACCCCGGAGTACCGGCGCGGCGGCGGCGGCGGCGGCGGCT 9148
QY 341 AlaAspArgAlaIleAlaLeuAspHisGlnGlyCysCysTYrLeuLeuAlaLeuAsp 360
DB 9149 GCGGACCGGCGCATGCGCTCTGACCAACGAGAAAGTGCTGCTGCTGCGGCTCGAC 9208
QY 361 ArgArgGlyHisAspAspGlyAlaArgAlaTrpLeuArgGluThrAlaGluThrLeuThr 380
DB 9209 CGCGGAGGCGACGACGAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9268
QY 381 GlyLeuAlaValArgAlaProAlaGluProThrProAlaMetValPheGlyIleProGlu 400
DB 9269 GGCCTGCGCTGCGCTGCGCTGCGCGGAGGCGGAGCGCGCGCATGCTTCCGAGGCTCCG 9328
QY 401 AlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHisAspValAspAlaTYrLeuLys 420
```



Db 9329 GCGGCGCGCGCTTCGCGCCCTGCTGCGCGACGCGACCAAGACGCC----- 9379  
Qy 421 Arg1LeaspGluCysLeuYsgLuiLeaArgAsnGlyGluSerYrrGluIleCysLeuThr 440  
Db 9380 -----TCGGCGCTCCGCAACGCGAGTCGTACGAGATCTGCTGACC 9421  
Qy 441 AsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuYrrSerAlaLeuArgAla 460  
Db 9422 AACATGCTCACCGCGCGACGAGCGACGCGCCCTGCGCTTACTCTCGCGCTGCGCGC 9481  
Qy 461 IleSerProValProYrrGlyAlaLeuLeuGluPheProGluLeuSerValIleuSerAla 480  
Db 9482 ATCAGCCCGCTCCGCTGCGCGCCCTGCTGAGTTCCTCGAGGCTGCTGCTGAGGCC 9541  
Qy 481 SerProGluLeuPheLeuThrIleGlyAlaAspGlyGlyValGluSerYrrProIleLys 500  
Db 9542 TCCCGCCAGCGGCTTCCTCAGATCGCGCGACGCGCGCTCGAGTCCAGGCCATCAG 9601  
Qy 501 GlyThrArgProArgGlyGlyThrAlaGluLeuAspGluArgLeuArgAlaAspLeuAla 520  
Db 9602 GGGACCGCGCCCGCGCGCGACCGCGGAGAGACGCGCTCCGCGCGACCTGCGC 9661  
Qy 521 GlyArgGluLeuAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAspLeu 540  
Db 9662 GCGCGGAGAGAGACCGCGCGAGAACCTGATGATCGTCACTGCTGCGCAAGACCTC 9721  
Qy 541 AsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGluThrTyr 560  
Db 9722 AACAGCTCTGCGGATCGGCTCGCTCCAGTCCGCGCTCTTCGAGTGGAGACCTC 9781  
Qy 561 AlaProValHisGluLeuValSerThrIleArgGlyArgLeuArgProGlyYrrSerThr 580  
Db 9782 GCGCGCTGCAACAGCTGCTGCTGACCATCGGAGAGCGCTGCGCGCGACCGAC 9841  
Qy 581 AlaAlaCysValAlaArgAlaAlaPheProGlyGlySerMetThrGlyAlaProGlyLysArg 600  
Db 9842 GCGCGCTGCTGCGCGCGCTTCCTCCGCGCGCTCCATGACCGCGCGCGCAAGAGCA 9901  
Qy 601 ThrMetGluIleIleAspArgLeuGluGlyProArgGlyValYrrSerGlyAlaLeu 620  
Db 9902 CCCATGAGATCATCGACCGCTGAGAGAGCGCGCGCGCTTACCGCGCGCGCTC 9961  
Qy 621 GlyTrpPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleValLeu 640  
Db 9962 GAGTGTTCCTCCCTCAGCGCGCGCGCGCGACCTCAGCATCGCTCCGACCATGCTG 10021  
Qy 641 AlaAspGlyGluAlaGluPheGlyValGlyAlaIleValSerLeuSerAspGlnGlu 660  
Db 10022 GCGGACGCGCGCGCGCGCTGCGGCTGCGCGCGCATCTGCTCCCTCGACAGGAG 10081  
Qy 661 GluGluPheThrGluThrValIleValAlaArgAlaMetValThrAlaLeuAspGlySer 680  
Db 10082 GAGAGCTTCAGGACAGCGCTGCTCAGGCGCGCGCATGCTCACCGCGCTCGACGCGAG 10141  
Qy 681 AlaValAlaGluAlaArg 686  
Db 10142 GCAGTGGCGGCGCGCGCA 10159

RESULT 6  
ARI98361 2220 bp DNA linear PAT 20-APR-2002  
LOCUS DEFINITION  
ACCESSION ARI98361 Sequence 14 from patent US 6352839.  
VERSION ARI98361.1 GI:20248210  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2220)  
AUTHORS Blanc,V., Thibaut,D., Bamas-Jacques,N., Blanche,F., Crouzet,J.,  
Barriere,J.-C., Debussche,L., Famechon,A., Paris,J.-M. and  
Dutric-Roseet,G.  
TITLE Streptogramins for preparing same by mutasynthesis

JOURNAL Patent: US 6352839-A 14 05-MAR-2002;  
FEATURES Location/Qualifiers  
source 1..2220  
/organism="unknown"  
/mol\_type="unassigned DNA"  
ORIGIN  
Alignment Scores:  
Pred. No.: 2,56e-80 Length: 2220  
Score: 1816.00 Matches: 385  
Percent Similarity: 64.03% Conservative: 76  
Best Local Similarity: 53.47% Mismatches: 207  
Query Match: 50.83% Indels: 52  
DB: 6 Gaps: 11  
US-10-089-514-2 (1-686) x ARI98361 (1-2220)  
Qy 1 MetArgThrLeuLeuIleAspAsnYrrAspSerPheThrHisAsnLeuPheGlnYrrIle 20  
Db 70 GTGCGAACCTCTGCTGATTCAGCAACTACGACTGTTCACTTCCAGATGCTG 129  
Qy 21 GlyGluAlaThrGlyGlnProProValValProAsnAsp-----AlaAspTrpSer 38  
Db 130 GCGGAGGTGAACGCGCGCGCTCGCTGCTGCTCGCAGACGACACCGCACCTGGCAG 189  
Qy 39 ArgLeuProValGluAspPheAspAlaIleValIleValSerProGlyYrrProAsp 58  
Db 190 GCGTGGCGCGCGCGCGCATTCAGCAACGCTGCTGCTCAACCGCGCGCGCGCGCGC 249  
Qy 59 ArgGluArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeu 78  
Db 250 ACCGACACCGACTGGAGCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 309  
Qy 79 GlyValCysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAla 98  
Db 310 GGGGTGCTGCTGGCGCACGAGCGCTGCTGCTGCTGCTGCGCGCGCGCGCGCGCGCA 369  
Qy 99 ProGluProMetHisGlyArgValIleSerGluValArgHisThrGlyGluAspValPheArg 118  
Db 370 CCGGAACCTTTTCAACGCGCGCGCACACGACATCCCGCACAGCGAGGCGCGTTCGCG 429  
Qy 119 GlyLeuProSerProPheThrAlaValArgYrrHisSerLeuAlaAlaThrAspLeuPro 138  
Db 430 AACATCTCTCTCCCGCTGACCGCTGCTCCCTTACCATCTCGTACCTCGCGCACTGCCC 489  
Qy 139 AspGluLeuGluProLeuAlaTrpSerAspAspGlyValIleMetGlyLeuArgHisArg 158  
Db 490 GCGGACCTGGGCGCACCGCGCACCGCGCGACGCGGACAGTATGCGCGCGCACCGCG 549  
Qy 159 GluLysProLeuTrpGlyValGlnPheHisProGluSerIleGlySerAspPheGlyArg 178  
Db 550 CACCTTCCCGCGCTTCGCGCTGCGATTCACCGCGCATTCATCGACGCGAACACGCGCAC 609  
Qy 179 GluIleMetAlaAsnPheArgAspLeuAlaLeu-----AlaHisHisArg----- 193  
Db 610 CGGATCTGCAACTTCGCGACGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCAC 669  
Qy 194 AlaArgArgHisGlyAlaAspSerPro----- 202  
Db 670 ACCGAAGCATACCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 729  
Qy 203 -----TyrGluLeuHisValAlaArgValAlaAspValLeuProAsp 215  
Db 730 GCGTCCCGCGCGCGTGGAGAGTACCGCGTGCATGTCGCGAGTCCCTGCGCGCGCAC 789  
Qy 216 AlaGluGluValArgArgGlyCysLeuProGlyGluGlyThrThrPheTrpLeuAspSer 235  
Db 790 GCGGACCGCGCGCTTACCGCGCTTCCGCGCACCGCGCGCGCGCGCGCGCGCTTGGCTGACAC 849  
Qy 236 SerSerValLeuGluGluValAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeu 255  
Db 850 AGCGCGTGGAGCGCGCGCTCGCGCGCTTACCTTCTGCGCGCGCGCGCGCGCGCGCGCTC 909



QY	256	AlaGluTYrLeuThrThyArgValAlaAspGlyValIleSerValArgGlySerAspGly	275
Db	910	GCGCAACAGATACCTTACGAGTCGCGCAACGGCGCGCGCTCAAG-----GACGCT	963
QY	276	ThrThrThrArgThrArgPro-----PhePheAsnTYrLeuGluGluLeuGlu	293
Db	964	TCAGCGCGGAGACCGCGCGCGCGCGGACCGCTCTTCGACCACTGGAAACAGAACTGGCC	1023
QY	294	ArgArgArgValProValAlaProGluLeuProPheGluPheAsnLeuGlyTYrValGly	313
Db	1024	GCGCCGCGCGCTGCCCC---GCCACCGCGCGCTGCGCTTCGAGTTCAACCTCGAGCTACGTCGCGC	1080
QY	314	TYrLeuGlyTYrTYrLeuLeuValAlaGluThrThrGlyAspProAlaHisArgSerProHis	333
Db	1081	TACCTCGGCTACGAAACCAAGGCGCAAGCGCGCGCGAGGAGCGCCACCGCGCGGAACTG	1140
QY	334	ProAspAlaAlaPheLeuPheAlaAspArgAlaIleAlaLeuAspHisGlnGluGlyCys	353
Db	1141	CCGCAAGCGCGCTTCATGTTGCGCGACCGGAGTCTCGGCTCTCGACCAAGAAAGGCGCG	1200
QY	354	CysTYrLeuLeuAlaLeuAsp-----ArgArgGlyHisAspAspGlyValArgVala	370
Db	1201	GCGTGGCTCTGCGCACTAGAGCAAGCAAGCGCGCGCGGACCGGACCGCGCGCGGAAAGCG	1260
QY	371	TYrLeuArgGluThrAlaGluThrLeuThrGlyLeuAlaValArg-----	385
Db	1261	TGGCTACCGAAGCGCGCGCGCGCGCTTCGCCAACACCGCGCGCGCGCGCTTCACCGTC	1320
QY	386	AlaProAlaGluProThrProAlaMetValPheGlyIleProGluAlaAlaGlyPhe	405
Db	1321	CTGCGCCGACGACCAACTGCGCGCTTGAC-----	1350
QY	406	GlyProLeuAlaArgAlaArgHisAspIysAspAlaTYrLeuIysArgIleAspGluCys	425
Db	1351	-----GTCCTACTACCGGCACAGCGCTGCGCGCGCTACCGGAAACGTGTGAGAAATGC	1401
QY	426	LeuIysGluIleArgAsnGlyGluSerTYrGluIleCysLeuThrAspMetValThrAla	445
Db	1402	CGCGCGCTGATCAACCGAGCGGAGACCTACAGAGGTGCTCGACGAACTCTCCGGGTG	1461
QY	446	ProThrGluAlaThrAlaLeuProLeuTYrSerAlaLeuArgAlaIleSerProValPro	465
Db	1462	CGCGCGCGGATGACCGCGCTACCGCGCTACCGCGCGCTGCGGACCGTAGCGCGCGCCC	1521
QY	466	TYrGlyAlaLeuLeuGluPheProGluLeuSerValIleuSerAlaSerProGluArgPhe	485
Db	1522	TACGCGCGCTACCTGACGTCCCGCGGGGCAACCGTGCTCACTCTTCAACCGGAAGGTTTC	1581
QY	486	LeuThrIleGlyAlaAspGlyGlyValGluSerIysProIleGlyTYrThrArgProArg	505
Db	1582	CTGCGCATATCGGCGCGAGGTGTGGCGGAGATCCAAACCATTCAGAGGACCGCGCGCGC	1641
QY	506	GlyGlyThrAlaGluGluAspGluArgLeuArgIleAspLeuAlaGlyArgGluIysAsp	525
Db	1642	GCGCGCGCGCGCGCGCAAGACGCGCGCGGTCAAGGCTCTCTCGCGCGCGCGAAGAGAC	1701
QY	526	ArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAspLeuAsnSerValCysAla	545
Db	1702	CGCACAGGAACCTGATGATGTGCACTTGTCGCGCAACGACCTCGCGCAGTCTGCGAC	1761
QY	546	IleGlySerValHisValProArgLeuPheGluValGluThrTYrAlaProValHisGln	565
Db	1762	ATCGGCTCTCGGTCCAGTACCGGCGCTGTTCAGAGTGAGAGACTTACGCGCACCGTCCACAG	1821
QY	566	LeuValSerThrIleArgGlyArgLeuArgProGlyIleSerThrAlaIleCysValArg	585
Db	1822	CTCGTCACACAGGTCGCGCGCGCTGCGCGCGCGACGTCCTCGCGCGCGCGCGGTACGG	1881
QY	586	AlaAlaPheProGlyGlySerMetThrGlyAlaProIlyslsArgThrMetGluIleIle	605
Db	1882	GCGCGCTTCCCGCGCGGTGATGACCGGCGCGCGCGCGCAAGGTCGACCATGCAATGTTATC	1941
QY	606	AspArgLeuGluGluGlyProArgGlyValTYrSerGlyAlaLeuGlyTYrPheAlaLeu	625

[illegible]



```
gene complement (2486 .3376)
CDS /gene="papC"
      /complement (2486 .3376)
      /gene="papC"
      /function="dehydrogenase"
      /codon_start=1
      /transl_table=11
      /product="PapC"
      /protein_id="AAC44867.1"
      /db_xref="GI:1575337"
      /translation="MRGGSYFRCVYVVGAGAVGRMFSHVLVASGVALVTMLDVAGCA
ADGSRVAGDVRKPGPEAVALAADVVLAPEPVAMEVEVLAGEVREGVALADTL
SVKSRLAGRLREAPGLQAVLPMFAPSLGIGRPVAAVVVDGCVRLVLVGM
GARVEMPARHDELTAQOQATHAALVAFGLIGELSDVVGALRDSAPPPLAMAL
LARIAGTPEVYFDIOANGAPAAARQALGRGLVRLGQAVERDEETPALFELMLGV
LGEHGALEBRLCARMTALH"
gene .33802
CDS /gene="papB"
      /gene="papB"
      /function="mutase"
      /codon_start=1
      /transl_table=11
      /product="PapB"
      /protein_id="AAC4868.1"
      /db_xref="GI:1575338"
      /translation="MTPPAIPAPAPATGPAAATPDIALRILDAADALDAVTRRL
DICTRIGEYKRLHQVPMOPHRIAQVHANARVAAHDGIDPAFLRTLITITETRL
EDEWIASGAPVPTPVHASASARGAVS"
gene .4677
CDS /gene="pamP"
      /gene="pamP"
      /function="N-methylase"
      /codon_start=1
      /transl_table=11
      /product="PamP"
      /protein_id="AAC4869.1"
      /db_xref="GI:1575339"
      /translation="MTAAAPTALQALDEATGQITGAGITADAAADPTLLAAHCOVA
PBDLDTCLAGPVPFRPMHYVRRRLREPERIVGHAYFMGHRDLAPGVFVPRPTEE
ITRDAIARLEALVRGTITAPLVLDLCAGGTAAVTLARIVPARVAGIEISQAAARA
RNRAGTGARIVOGDADAPELSGVTVTPRYPIGLRTSAREVEHDPRLAM
AGEEGIGMIKAMERTAAIRLLAPGVLLLEHGSYQLASVAPLPRATGRMHSASRPTCN
DGLTRVNRHTCAPA"

ORIGIN

Alignment Scores:
Pred. No.: 5.86e-80 Length: 4740
Score: 1016.00 Matches: 385
Percent Similarity: 64.03% Conservative: 76
Best Local Similarity: 53.47% Mismatches: 207
Query Match: 50.83% Indels: 52
DB: 1 Gaps: 11

US-10-089-514-2 (1-686) x SPUB0417 (1-4740)

QY 1 MetArgThrLeuLeuIleAspAsnTYrAspSerPheThrHisAsnLeuPheGlnTYrIle 20
   :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 77 GTCGGAACCTGCTGATCGACACTACGACTCGTTCACTACCAACCTCTCCAGATGCTG 136

QY 21 GlyGluAlaThrGlyLeuProPheValValPheAsnAsp-----AlaAspTrpSer 38
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 137 GCCGAGGTGAACCGCCCGCTCGCTCGTCGCGAAGCAGCACACCGCGACCTGGCAG 196

QY 39 ArgLeuProValGluAspPheAspAlaIleValIleValSerProGlyProGlySerProAsp 58
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 197 GCCCTGCGCGCGGCGCACTTCGACAAAGCTGCTCTCACCCCGGCGCCACCCCGCC 256

QY 59 ArgGluArgAspPheGlyIleSerArgAlaIleThrAspSerGlyLeuProValLeu 78
   :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 257 ACCGACACGCACTGGGCGCTCAGCCCGGGGTGATCATCGAATGGAGACCTGGCGCTGCTC 316
```

```
QY 79 GlyValCysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAla 98
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 317 GGGGTGCTGCTGGGCCACCGCCCTGTGCTGTGCGCGCGCGCGCGCTGTCACCGCA 376

QY 99 ProGluProMetHisGlyArgValSerGluValArgHisThrGlyGluAspValPheArg 118
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 377 CCCGAACCTTTTCACGCGCGCACCGACGACATCCGCGCACGACGCGGCGCGCTTCCGCG 436

QY 119 GlyLeuProSerProPheThrAlaValArgTYrHisSerLeuAlaIleThrAspLeuPro 138
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 437 AACATCCCTCCCTCCGCTGACCGGTGTCGCTGACCTGCTGACCTGCGGCAACCTGCC 496

QY 139 AspGlyLeuGluProLeuAlaTrpSerAspAspGlyValValMetGlyLeuArgHisArg 158
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 497 GCCGACCTGGCGGCCACCGGCCACACCGCGGACGGGACAGTATGGCCGCGCCACCGCC 556

QY 159 GlyLeuProLeuTrpGlyValGlnPheHisProGlySerIleGlySerAspPheGlyArg 178
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 557 CACCTGCCCGCTTCGCGCTGCAGTTCACACCGGAATCATCAGCAGCGCAACGCGCAC 616

QY 179 GluIleMetAlaAsnPheArgAspLeuAlaLeu-----AlaHisHisArg----- 193
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 617 CGGATCTCGCCCAACTTCGCGGACCTGTCCCTGCGCGCGCGCGCGCACCGCCCGCGCAC 676

QY 194 AlaArgArgHisGlyValAlaAspSerPro----- 202
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 677 ACCGAACGATACCCGCCACCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 736

QY 203 -----TYrGluLeuHisValArgArgValAspValLeuProAsp 215
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 737 GCGTCGCGCGCGCGGTGGGGAGTACCGCGCTGAGTGTGCGGAGGTGCGCTGCGCGCGCAC 796

QY 216 AlaGluGluValArgArgGlyCysLeuProGlyGlyGlyThrThrPheTrpLeuAspSer 235
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 797 GCGGACGCGCGCTTCAACCGCGCTGTGCGCGGACCGCGCGCGCGCGCTTCTGCGCTGACAG 856

QY 236 SerSerValLeuGluGlyValAspArgPheSerPheLeuGlyAspAspArgGlyProLeu 255
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 857 AGCCGCTGAGCGCGGCGCTCGCGCTGACCTTCTGAGGCGCGCGCGCGCGCGCGCTC 916

QY 256 AlaGlyTrpLeuThrTYrArgValAlaAspGlyValValSerValArgGlySerAspGly 275
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 917 GGGGAACAGATACCTACGACGCTGCGCGACCGCGCGCGCTGCGGCTCAAG-----GACGGT 970

QY 276 ThrThrThrArgThrArgArgPro-----PhePheAsnTYrLeuGlnGluLeuGlu 293
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 971 TCAGCGCGGAGACCGCGCGCGCGCGCACCTTCTTGCACCTGGAACACCGAACCTGGCC 1030

QY 294 ArgArgArgValProValAlaProGluLeuProPheGlnPheAsnLeuGlyTYrValGly 313
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1031 GCCCGGCGCGCTGCC---GCCACCGGCGCTGCGCTTGAAGTTCAACCTGCGCTACGTCGGC 1087

QY 314 TYrLeuGlyTYrGlyLeuLeuAspAlaGlnTrpGlyAspProAlaHisArgSerProHis 333
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1088 TACCTGCGCTACGACACCAAGCGCCACAGCGCGCGGAGAGACGCCACCGCGCGGCAACCTG 1147

QY 334 ProAspAlaAlaPheLeuPheAlaAspArgAlaIleAlaLeuAspHisGlnGlnGlyCys 353
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1148 CCCGACGCGCGCTTCATGTTGCGCGACCGGAGTGTGCTGCGCTGACACGACGAGGCGCG 1207

QY 354 CysTYrLeuLeuAlaLeuAsp-----ArgArgGlyHisAspAspGlyValAlaGlyAla 370
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1208 GCGTGGCTCTGCGACCTGAGCAGACACCGACGCGCGCGCGCACCGCGCGGCAACCGC 1267

QY 371 TrpLeuArgGluThrAlaGluThrLeuThrGlyLeuAlaValArg----- 385
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1268 TGGCTACCGGACGCGCGCGGACCTTCGCGCACACCGCGCGCGCGCGCGCTTACCTG 1327

QY 386 AlaProAlaGluProThrProAlaMetValPheGlyIleProGluAlaAlaGlyPhe 405
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1328 CTGCGCGACGACCACTGCGCGCGCTGAGC----- 1357

QY 406 GlyProLeuAlaArgAlaArgHisAspArgAspAlaTYrLeuLeuArgIleAspGlyCys 425
```



Db	1358	-----GTCACCTACCGCCGACGCTGCCCCGCTACCGGAACTGTGCGAGATGC	1408
Qy	426	LeuLyGluIleArgAnGlyIleSerThrGluIleCysLeuThrAsnMetValThrAla	445
Db	1409	CGCGCCCTGATACCAAGCGGAGACCTTACGAGTGTGCTGACGAACATGCTCCGGGTG	1468
Qy	446	ProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuArgAlaIleSerProValPro	465
Db	1469	CCGCGCGGATCGACCGCTCACCGGCTTACCGGCGCTTGCGGACCGTCAAGCCCGCCCC	1528
Qy	466	TyrGlyAlaLeuLeuGluIlePheProGluLeuSerValLeuSerAlaSerProGluArgPhe	485
Db	1529	TACGCGCCCTACCTGAGTTCGCCGGGCGCACCGTGTGCTGACCTCCACCGGACGTTTC	1588
Qy	486	LeuThrTlleGlyAlaAspGlyIleGluSerTyrProIleTyrSerThrArgProArg	505
Db	1589	CTGCGCATCGCGCGGAGGTTGGCGGAGTCCAAACCATCAAGGACACCGCCCGCCGC	1648
Qy	506	GlyGlyThrAlaGluGluAspGluArgLeuArgAlaAspLeuAlaGlyArgGluTyrAsp	525
Db	1649	GCGCGCGCGCCCGCCAGAGCGCGCGCTGCAAGCGCTCCCTCGCGCGCGCGGAGGAGC	1708
Qy	526	ArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAspLeuAsnSerValCysAla	545
Db	1709	CGAGCGAGAACCTGATGATCGACCTGCTCGGACCACTCGGCGAGCTCTCGGAC	1768
Qy	546	IleGlySerValHisValProArgLeuPheGluValGluThrTyrAlaProValHisGln	565
Db	1769	ATGAGCTCCGTCCAGTACCGGCGCTGTTGAGTGGAGACCTTACCGCCGCTCCACG	1828
Qy	566	LeuValSerThrTlleArgGlyArgLeuArgProGlyThrSerThrAlaAlaCysValArg	585
Db	1829	CTCGTCAGACCGGTCCGCGCGCGCTGCGCGCGGAGCTCCGCGCGCGCGGAGTACG	1888
Qy	586	AlaAlaPheProGlyGlySerMetThrGlyAlaProLysLysArgThrMetGluIle	605
Db	1889	GCGCGCTTCCCGCGCGGTGATGACCGGCGCGCCCAAGCTCCGACATGCGAGTTCATC	1948
Qy	606	AspArgLeuGluGluGluProArgGlyValTyrSerGlyValAlaLeuGluIlePheAlaLeu	625
Db	1949	GACCGGCTCGAGAAAGGCGCGCGGTGCTGCTGCGCGCTGCGGCTGCTGCGGCTC	2008
Qy	626	SerGlyAlaAlaAspLeuSerIleValIleArgThrIleValLeuAlaAspGlyGlnAla	645
Db	2009	AGCGGCGCGCGCGACCTCGACATGCTCATCGCACCATGCTGCCACCGAGGAGCGCGC	2068
Qy	646	GluPheGlyValGlyGlyAlaIleValSerLeuSerAspGlnGluGluPheThrGlu	665
Db	2069	ACCATCGGCGTGGCGCGCGCTGCTGCTGCTGCGACCCGACGACGAGGTCCCGGAA	2128
Qy	666	ThrValValIleAlaArgAlaMetValThrAlaLeuAspGlySerAlaValAlaGlyAla	685
Db	2129	ATGCTCTCAAGCGGACACCTGCGCGCGCTG--CGCAGGACACAGCGGCGCGCC	2185
RESULT 8			
STMPABA			
LOCUS	4607 bp	DNA	linear BCT 31-AUG-1993
DEFINITION	Streptomyces griseus pab gene fragment.		
ACCESSION	M93058		
VERSION	M93058.1	GI:153396	
KEYWORDS	pab gene.		
SOURCE	Streptomyces griseus		
ORGANISM	Streptomyces griseus		
REFERENCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
AUTHORS	Criado,L.M., Martin,J.F. and Gil,J.A.		
TITLE	The pab gene of Streptomyces griseus, encoding p-aminobenzoic acid synthase, is located between genes possibly involved in candididin biosynthesis		
JOURNAL	Gene 126 (1), 135-139 (1993)		
MEDLINE	93231527		

	PUBMED	8472954	
COMMENT	Original		
FEATURES	Location/Qualifiers		
source	1..4607		
	/organism="Streptomyces griseus"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:1911"		
	/cruise_1fb="IMR03570"		
	88..1173		
CDS	/note="thioesterase-II-S-acyl fatty acid synthase homolog"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="AAA72110.1"		
	/db_xref="GI:388262"		
	/translation="MRVYVDSQCVAGGCVLNADPEVPDQDGVVLLRADPTSGTT		
	RSARATCARPRSSRRTERPAGADRRCPVRGPRTRGAPADPPATDRONHR		
	DSPVTTADDTAARMLRRHPRADAVRLVCFPHAGSGASFHYPRARPRAGAVVSLQ		
	YFGRDRRKEPCVPLDGLTADLITGLDLERPTVFVFGSHGALALAFETVRLKQ		
	AGPRVIAVSGRRGPSTTRAEVHTRDDDDGIVAEKRLGTAAVGLGDEILMALPAL		
	RQDYVAIETVTCPPDRRLACGLTVLTGEDDPLTVEEAEKMRDHTTGPRRLRVFTGGH		
	FPLTGLDVAVNTIEIAQLHPRAAPAA"		
	1370..3541		
gene	/gene="pabAB"		
	1370..3541		
CDS	/gene="pabAB"		
	1370..3541		
	/transl_table=11		
	/codon_start=1		
	/product="p-aminobenzoic acid synthase"		
	/protein_id="AAA72111.1"		
	/db_xref="GI:388263"		
	/translation="MRVTLVNDYDSFTYNLPHYLSRANGREBEVIRNDPAMRGILLD		
	APDNLVSPGPTPHRPADFGICARIAEGRILPVIGVCIHQGMALAHGARVRAPRP		
	RGRISAVHSDGTGFEGLPEPLVVRVYSGLVLTLPPLATAMSDDGLMALRHRT		
	LPLKGVQHPHSEIGTODGHRLLANFRDLTERHGRTRHGRAGHTLPPAPAREKAT		
	TCTPRRLRIASLPTRMDAARVPSLPRTGHPHWLDSRRSGEGLGGLSMGDSAP		
	LARTAKADVAHTVTVRADGASTVESAPLTLVENDLAKRTVEPELPFARALGWGC		
	LGEYLKACDGDAAHRSBPDAVLVFAFARVALDRTRTYLLALVEDDAEAEARWL		
	LAASATLDDAVAREPEPCPEAPVCTGPEVLEHDBDGYLKLIDVCOQEIAGETVEVC		
	LNNMAEDTDLTPMAAYRALRRVSPAPFAFLDFGPMVLSSEPFLRIDHGRMS		
	RKIKGRPRGATPOEDALVRLATCEKORAEINMIVDIVRHDLRCRCAVSGVMDPV		
	FOVEYATYTHQIVSTVTARLREDSFPVAARAPFAGSGTTCAPKRTMOTIDRLBEGP		
	KQVYGAIGATFSLTGAVIDSLVIRTVISGGGLRVGGGAVLALSDPADEFETAVKA		
	APLRLDLTAPFPRRRPKDLIDGDDDDGDAAPKDLVLPG"		
	3685..>4607		
	/note="p-coumarate:CoA ligase homolog"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="AAA72112.1"		
	/db_xref="GI:388264"		
	/translation="WVPVHADYVTDPSPTGRTLDGLTPRVFADAVHGGDAVALV		
	DGEVALTWASAMRTAVDALARGDSGVSDVVALHLPMSWEYTLHLAAASVGAVTM		
	PHQGNAPSDVALLERVPRAVVLTAQOEGGGLTGPALEVLPELRAYLVTDGAA		
	GEGTETVTEMLRMSGEDPLVYEVPRDSFLLPSSGTTTSARPKTCLSHHEGLNRSR		
	AATEETADAVAGTLITACPLTHCFGLQSYSLFPAGRVLLSGMDVGFLELARRR		
	PSVVAVVPASCTTWSPGCAKRTATAPASRPD"		
ORIGIN			
Alignment Scores:			
Pred. No.:	1,66-77	Length:	4607
Score:	1766.00	Matches:	379
Percent Similarity:	63.88%	Conservative:	72
Best Local Similarity:	53.68%	Mismatches:	211
Query Match:	49.43%	Indels:	44
DB:	1	Gaps:	12
US-10-089-514-2 (1-686) x STMPABA (1-4607)			
Qy	1	MeArgThrLeuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGlnTyrIle	20
Db	1370	ATGCGGACCCCTTCTGTCGACAACTHACGACTGTTCACTTAACAACCTTCTTCACTHACCTC	1429
Qy	21	GlyGluAlaThrGlyGlnProProValValProAsn--AspAlaAspTyrSerArg	39







gene; canf gene; canfA gene; canf gene; Cho-like protein;  
cytochrome P-450; ferredoxine; PABA synthase; thioesterase.  
SOURCE  
Streptomyces griseus  
ORGANISM  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomyicinae; Streptomycetaceae; Streptomyces.  
REFERENCE  
1  
Campelo, A.B.  
Thesis (2000) Department of Microbiologia, Universidad de Leon,  
Leon, Spain  
AUTHORS  
2  
Campelo, A.B. and Gil, J.A.  
REFERENCES  
The candididin gene cluster from Streptomyces griseus IMRU 3570  
JOURNAL  
Microbiology (Reading, Engl.) 148 (Pt 1), 51-59 (2002)  
MEDLINE  
21642576  
PUBMED  
11782498  
REFERENCE  
3 (bases 1 to 39314)  
Gil, J.A.  
Direct Submission  
Submitted (14-NOV-2000) Gil J.A., Microbiologia, Universidad de  
Leon, Campus de Vegazana s/n, 24071, SPAIN  
FEATURES  
source  
location/Qualifiers  
1. 39314  
/organism="Streptomyces griseus"  
/mol\_type="genomic DNA"  
/strain="IMRU 3570"  
/db\_xref="taxon:1911"  
complement(<1..1785)  
/note="ORF1"  
/codon\_start=1  
/transl\_table=11  
/product="Cho-like protein"  
/protein\_id="CAC22112.1"  
/db\_xref="GI:12231154"  
/db\_xref="UniProt/TREMBL:O9EWC6"  
/translation="MLBERTETALVDALGAAITGRSSIVLTITGLIGRSALQRL  
SCLADRDVRAVIRAHAAPEODFAVGVRQLPETLLGSDPEDARDMDAHAFARH  
VLADDAIPGADQAIATEAVLHGLSLANYSADSRLLILVDDLQMSVPSIRMLTY  
LAKRLHRAVVCALRGDPSRSHHTLVREIRPAATOTLRPASLSLTATHELVREHFG  
EAGDDPFVQACHASGNPLILSTIIWGTCGRPLAAHAEARLRRLRSQLEERLAS  
ILRTQAPAPRDLLAAITAIIGESQDAPTLRLALDSTIGAGALRALGALAAPEP  
RFTHRSVRDAESTLTWQKERHDEPAALVYAGSPAEQVAAQLMAVTRQPMADV  
VIRAAADTLARRGAPDTAGYLCRALDSPAAGVGRLLVELGTAERGDPACBRH  
IAQMTLLPEPRDAVAALRIIPTALGPALITVDLLRQAEDLGPADGIGTDRDLA  
LRLEARLRHCGHEDPHELAASARHLKEIGPEPVDGTGERELTIALGATTCGLHPA  
SEVYRLARGVLDSEPAHSRIHSTLPLAT"  
9525..10583  
/gene="canA"  
9525..10583  
/gene="canA"  
/codon\_start=1  
/transl\_table=11  
/product="aminotransferase"  
/protein\_id="CAC22113.1"  
/db\_xref="GI:12231155"  
/db\_xref="GOA:O9EWC5"  
/db\_xref="UniProt/TREMBL:O9EWC5"  
/translation="MAFTHVSRPALDGRLEAVSVASGSGTSSOGPVYRPEEFA  
ENGVAHGVAACSSGTALTLALRALITGREDETIVEPFTWVASAMATTGATPVVD  
CGDLDNIIVRIIEKITPTTRAVMPVHVGRCDMAVMDLAQYMLRVVEDSAEAG  
VRPVDIACPSLPAFKITITAGEGVCLTDPRLABQALRAMAFRVEDSFLKKAAY  
NYRMTMOGAVALAOETERLDEILATRREIEARVDAGIKDLPGITLMPARDVLMYDLR  
AERREELRAHLDERGIETRLFFKPMRQPEYLDPMVPTLNAHRFSDGLYLPHTGLT  
AAQOEYITGAVRPYRAG"

/db\_xref="GOA:O9EWC4"  
/db\_xref="UniProt/TREMBL:O9EWC4"  
/translation="MTSPGPTVVDVFPRTTPEPPLQSYAEHRKONGLVQTHLPNGR  
PIWLVTRHEDVRAVLTNPRISANPNDEGPGVGMWGKQOIGMFGVGLSPEDR  
PRKVLIPERTVRAVVELARPAIRTYDEPDIAMLAGNTADIVNDALPVPISVIAL  
GPSADRDFFESRTITLVARTSTDEERAEATROLRLRTINRLIVYKKKRGSDILSR  
LSTGKLSDEESGLVLLLLIAGHETTANNIGAGVVTLLSHREWIGDDRIVEBLRHS  
VADWALRYAVADVDEIAGQITRKGGIYVPLLASAMHDEAFQCPAPFNPETERHVA  
FGYGHQICIGQNLVWEMEIAVRKLFERIPEIRLAVPEQQLAKYKDGILLFGHLEPVR  
W"  
11906..12100  
/gene="canP"  
11906..12100  
/gene="canP"  
/codon\_start=1  
/transl\_table=11  
/product="ferredoxine"  
/protein\_id="CAC22115.1"  
/db\_xref="GI:12231157"  
/db\_xref="GOA:O9EWC3"  
/db\_xref="UniProt/TREMBL:O9EWC3"  
/translation="MRVTYDSQGCVGAGQCVLNADPEVPQDDGVVLLRAEPDRDH  
EAVRTAGDLCPSASVVLQED"  
12223..12993  
/gene="canT"  
12223..12993  
/gene="canT"  
/codon\_start=1  
/transl\_table=11  
/product="thioesterase"  
/protein\_id="CAC22116.1"  
/db\_xref="GI:12231158"  
/db\_xref="GOA:O9EWC2"  
/db\_xref="UniProt/TREMBL:O9EWC2"  
/translation="MTTADDTAARLNRHPRHDAVNLVCPERHAGSASFYHYSAR  
FAPGAVSVSLQYPRQDRKEPCVDELGLADLITLQQLPLDERPTVFGHMGALAA  
FPTGARLBQKGPRTVIAISGRGSPVIAERAVHTRDDGIVAEKRLNGTGAAGVGD  
EELILMALPALGELDRALETYTCPDRIACGLTVLTGDDPLTVEEAERWRDHTTG  
PRLRVFTGHEFFLTQHLDAVNTETIAQALHPDAAPAA"  
13190..15361  
/gene="pabAB"  
13190..15361  
/gene="pabAB"  
/codon\_start=1  
/transl\_table=11  
/product="PABA synthase"  
/protein\_id="CAC22117.1"  
/db\_xref="GI:12231159"  
/db\_xref="GOA:O9EWC1"  
/db\_xref="UniProt/TREMBL:O9EWC1"  
/translation="MRTLLVDYDSFTYVLPHYLSRANGREPEVIRINDDPAPWPGILD  
APDNVLSRPGCTGPHRPADFGICARIAERGLPVIGVCLGHQGMALHAGARVRAPEP  
RNGRISAVAHNDTGLEPGILPOPLEYVRHSLAVTELPEPLEXTANSSEDCVAMALNRT  
LPLWGVQHPRESIGTQDGRHLANFRDLTERGRTRHGAGHGLPPAPARETKAT  
TGTPRLRLVIANKSLPTWMDAEVAFDSLFRFGHPMLDLSRSGELGQLSMWGDASGP  
LARTAKADVAGTIVTRADGASITVESAFILTWLENDLAGLRTVEBELPPAFALGWGVC  
LGYELKAECDGGAARSPDDAVLVFADRALVDHRTTYYLALVEDDABAEAAEMV  
LAASVTLDEADGAAHRSPEDEAPVCTGPELPHDDGYKXILDVCOETIAAETAEVC  
LTNMBADDTLTPMAAYRLRVSAPPAFLDPGMAVLTSSPERFLRIDHGMES  
KIKTGRPRGAPQEDALVRLATCEKDRANLMTVIDLRHLCRCAVGSVAVDVP  
RVEYTAIVHQVSVTVTARLREDSVPVAVARAFQGSWTGAPKIRTMQIIDRLBGP  
RNVYGAIAIFYELTGAVIDLSIVIRVYVSGSRLRYGVGGAVALSDPADEFETA VKA  
ARLLLDPAFPGREAPKGDLDGEEDDGDADAAPDLVLPG"  
20829..21836  
/gene="canRA"  
20829..21836  
/gene="canRA"  
/codon\_start=1  
/transl\_table=11  
/product="ABC-transporter"  
/protein\_id="CAC22118.1"  
/db\_xref="GI:12231160"  
/db\_xref="GOA:O9EWC0"







Db 14762 GCCCGCTCTAGTGGCTCCCTGGCCACCTGGCAGAAAGGACCGCCCGAAGAACTGATGATC 14821  
Qy 533 ValAspLeuValArgAsnAspLeuAsnSerValCysAlaIleGlySerValHisValPro 552  
Db 14822 GTGACCTGGTCCGCCACGACCTGGGGCGGTGGCGGAGGTGGCTCGTCTGGTGGCCAC 14881  
Qy 553 ArgLeuPheGluValGluThrTyrAlaProValHisGlnLeuValSerThrIleArgGly 572  
Db 14882 CCGGTGTTCCAGGTCGAGACGTACGCGCACCGCTCACGCTGCTCAGTACCGTCACGCGC 14941  
Qy 573 ArgLeuArgProGlyIleThrSerThrAlaAlaCysValArgAlaIlePheProGlyGlySer 592  
Db 14942 CGCGTGGCCAGACACGACCGCGGTGGCGCGGTGGCGGCTTCCTCCCGCGGGGTG 15001  
Qy 593 MetThrGlyAlaProIlySlyArgThrMetGluIleIleAspArgLeuGluGluGlyPro 612  
Db 15002 ATGACCGGGCGCGCGAAGATCCGACCATCATATCATGACCGGTGGAAGCGCGGCC 15061  
Qy 613 ArgGlyValTyrSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSer 632  
Db 15062 CGCGGTCTTACTCTCGGGCGCATCGGCTACTCTCTCCCTCACCGCGCGGTAGACTGTCC 15121  
Qy 633 IleValIleArgThrIleValLeuAlaAspGlyAlaGluPheGlyValGlyGlyAla 652  
Db 15122 ATGTGATCCGACGGGTGGTCTGACGGCGGCGAGCGTCTGCGGTGCGGCGGCGC 15181  
Qy 653 IleValSerLeuSerArgGlnGluGluPheThrGluThrValValIlyAlaArgAla 672  
Db 15182 GTGATGCGGCTCTCCGACCGGCGCGACGAGTTGAGAGAGCGGCGGTCAAGGCGGCCG 15241  
Qy 673 MetValThrAlaLeuAsp 678  
Db 15242 CTGCTGCGTCTCTCGAC 15259

RESULT 10  
AY310323  
LOCUS 138203 bp DNA linear BCT 06-DEC-2003  
DEFINITION Streptomyces sp. FR-008 heptane macroide complex synthesis gene  
ACCESSION AY310323  
VERSION AY310323.1 GI:34766435  
KEYWORDS  
SOURCE Streptomyces sp. FR-008  
ORGANISM Streptomyces sp. FR-008  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 138203)  
Chen, S., Huang, X., Zhou, X., Bai, L., He, J., Jeong, K. J., Lee, S. Y. and  
Deng, Z.  
Organizational and Mutational Analysis of a Complete  
FR-008/Candididin Gene Cluster Encoding a Structurally Related  
Polyene Complex  
Chem. Biol. 10 (11), 1065-1076 (2003)  
14652074  
2 (bases 1 to 138203)  
Chen, S., Huang, X., Zhou, X., He, J., Bai, L., Jeong, K. J., Lee, S. Y. and  
Deng, Z.  
Direct Submission  
Shanghai (30-MAY-2003) Bio-X life Science Research Center,  
Shanghai Jiaotong University, 1954 Huashan Road, Shanghai 200030,  
China

FEATURES  
source 1. .138203  
/organism="Streptomyces sp. FR-008"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:206662"  
/complement(574. .1950)  
/gene="fscO"  
complement(574. .1950)  
/gene="fscO"  
/note="putative FAD-dependent monooxygenase"  
/codon\_start=1  
/transl\_table=11

gene  
CDS

/product="FscO"  
/protein\_id="AA082549.1"  
/db\_xref="GI:34766436"  
/translation="MPSTCAPTSGFPGCPHRRKGRKRSRIAVPTPCCGKRGH  
CHAKKRVSMRGGSVAIVGSGITAGCAALAAARGAERTVTERDARLRDGVGIG  
HPARYAELRDAGYADDEMPALNRNRYVSRGDSEHRAITGTQPPPPAYSGILMS  
ELRRVPEPTATIRAGAKVEAEODDAGVTALDGPPEPFDVITGADGRSVRELMY  
PGTGAAYAGYIGMRGTSPDVSGLPDGDANHITPEPGCQMYRI.PDSGGRLMWL  
YTPAPRIDGLHEDLTPTSLPEGRINAEITELRALVAHEHPPFPAAKLSTPAETTF  
IOPIDVDVPHYATDRMYLIGDAVSARPHLDAGSVKALQDATALAAMVAGESWKEI  
LEGHYAARGPVTAVGLARRNGSAQVETTPWSAMQAEFPAMMQEONNGSDRSGF  
GCHSLKSR"  
complement(2264. .3037)  
/gene="pabC"  
complement(2264. .3037)  
/gene="pabC"  
/note="putative 4-amin-4-deoxychorismate lyase"  
/codon\_start=1  
/transl\_table=11  
/product="PabC"  
/protein\_id="AA082550.1"  
/db\_xref="GI:34766437"  
/translation="MIEIDGEPAGEALASALTNVGHFTTLVNGRYRGV.DLHLER  
LIRDCRTPDALDPDRVKLARARAPDGRATVYVYDPALNIGNIAADARBGILV  
TSRPADKPEGLRVSVYHRDLPEVSVGCPYLRLRQKORAGYDVLFTGGDG  
ILEGTMVNGVLDDEVWVGGEVLATGTRQLRGHRSHRTGPRGRLSRGGILD  
QRGVVRPVYTGIDREFPAAHSHSVRLAEIYALGSPPL"  
complement(3150. .3818)  
/gene="fscR1"  
complement(3150. .3818)  
/gene="fscR1"  
/note="putative transcriptional activator"  
/codon\_start=1  
/transl\_table=11  
/product="FscR1"  
/protein\_id="AA082551.1"  
/db\_xref="GI:34766438"  
/translation="MDPAPAADPAGSPERRAAVLADGPDRAAGYVACLDPSLTIOVNO  
EERERFGSSSELCSRFCDLVHPVOQPLMHQFAPMLDKGRATEVIAVDOERTA  
STLPNALAARGRTPDVAIIVWMAAEAGDVMAPRKLLSEIDARILLEGIAA  
GVSTIPLASRLYLSRQGVHYHTGLRKLVNRAALVSRAVSMGLKVTMPKPYVE  
DPKX"  
complement(4377. .7205)  
/gene="fscR11"  
complement(4377. .7205)  
/gene="fscR11"  
/note="putative transcriptional regulator"  
/codon\_start=1  
/transl\_table=11  
/product="FscR11"  
/protein\_id="AA082552.1"  
/db\_xref="GI:34766439"  
/translation="MLERTETIALVDAALGAATGSRSLVLTGPIGIGRSALORL  
SCLADRDVRYLRHAAPMEDQPAFCVYRQIFPTLLGDSPEADADRMDADAAPRKH  
VLADDAAPGADQAIATEAVLHGLSLIYASVSRLILVDDIQMSVPSRLWITY  
LAKRLHGLAAVVVCALRDDPSRSHLTAREIEAQTQTLRLPISLTATNHELVRHFG  
EAGDPEYVQACHASVGNPLFLISLITGFGIRBPDLAEHARTARLPPSGRRLAS  
IKRQDPAPVDAIAAIVLIGEOSDAPTLARLAGDSIGVAGALRALGALGAAPREP  
RRIHSVYPAESTLTMOEREMHDEAALIVAAASPVOVAAQIMAAVTRPOPAWV  
VIRAAADTLRRGADPTANGYICRALDSPPAAGVGRGLVLTGTRBGFDPDLASRH  
IAQAMTLLPEPRDAVVAALRISPTALGAPVLAVDLIRQAADLGPADSLTQDOLA  
LRLERLRHCGHEDPHELAASARLKEIGEPVPVDTGGERLTIALLGATVGCILPA  
SEVVLACGLVDREPTASRIHSTPLAITTIAADSVASVSMILAEHARGCHLPA  
ANALVHIBQALVHLRGRLAORBOSETLTGWTETPAHQGLDTPAATMTAIVLHNSDP  
SLAEKYLRRASRIROGLTITGLILLETRETRIAAGDQALGALSICGRCOLEAGCMH  
NFVLFPWRPAVGVHRRLGDIITARSLEBEYARABOWKAPVGVGRLALRLRLEDG  
RGILGLREAVTLRGSANOLELARAISTLAELGGGAETELAREATLATISCGAAGL  
AERARAGHRSGPATPEVDVLTPTERRVGLACRGLTIOETIAGALGVSSRAVREKLT  
HAYRRLGISGRELIVLFSRAEYAGCD"  
complement(7210. .10254)  
/gene="fscR11T"  
complement(7210. .10254)  
/gene="fscR11T"  
/gene="fscR11T"

gene  
CDS



```
/note="putative transcriptional regulator"
/codon_start=1
/translate=1
/product="fscR11"
/protein_id="AA082553.1"
/db_xref="GI:34766440"
/translation="MYDLHMPLEPLEERRELAASAVIGELSGRAVAVVTGEP
GLGNDLIRMAAAYATDAGLVLSAHTPAHEHYRQVVAQLAEENRALAPFLFD
EQQGLPEGLNGLNVAVDRLVRLVDRDOWLDPASLRLMALRDLPRALVLTST
GTALTREMSVGTPLTGLATTTELALPLTSGGTATVILARAGDPATTEALERT
RGI.PAVVHDVLDPRARAGYSPRRDRIDPLALTHEVGDHATRLSDLRPALVDAR
ALAVCGDLDFPLVCTLAGHSVSESLRALALASGLTTLRDHPVQDVFARVLE
EMPAADAEIYARAAQGLAQRVAADDOGILILARVGPMAVDILRQDVTALVGE
RDLAVAYLARLADDEPLAEDRARIEFOLASVENVTPTAERLIGLILATPGGA
GLBARATDLCIGDDTPRAARHALAGATISAPAPRPPRGPTGA.PAVPEPESTPT
HLYPHRYACTPGGATGPDGRSAPRLPCTPRAHADPRS.PGPAAGRPPTAGDEP
AVABHELVLCVASFLRHDDADLDVLPVLPDVAHPAVAGVMAWEGVAVNHE
EVRLARSAFVPATGRPLLVPRLLGCRALLTDDDEENHGLALLDSHERAI
SVAILTVRAELHIRHGRPDAAARDLAAQAEPLDLHPLFPLYMALSMITDQNG
HTRARATARPPLPLAESATTAOLIFAVGLARTDDEHOBREHRCAGMLLHG
CANPAVOPWRSILAAEAHTGDTREARIVHEVRILARWGGAAPLGRADSLAVTE
ENRVENDIRAVTTLASASARTATYTRAVIETLAAERENERTVAGDPATVLPVAP
PGPHVSAVAPRPPGPAVLGTGRITAKAGHDAVAWHGLTAVERTALAAQGLNR
EIALTELAVTTRAVELRISGVYRKIRIRGREBELRALVOEAGS"
/complement(10298..13315)
/gene="fscR1V"
/complement(10298..13315)
/gene="fscR1V"
/note="putative transcriptional regulator"
/codon_start=1
/translate=1
/product="fscR1V"
/protein_id="AA082554.1"
/db_xref="GI:34766441"
/translation="MMLRWAGDRSGAAPRSGSAGPPSLAPDARES.PVPHRSAP
PAPPLVGRHQLDVLARHADARARPRVLLDGPAGICTALIRALADGCLAGT
TLYGSCRAVAVATGSGVRALFGALGLTGKRTSPILVCGARAPLALADGDEDA
DPGSTFVLOGLVLANLMAADGVLVLDVWCDERSLRMIDFLIRAPDPLILVY
AARSTGCLTAPDALADVAHNL.PASIGLGRPLTAAYAEITANSPEODPLSEVRL
LSTVGSPLLEIVRLDLDRSAGLPDDTGARLQAQGVAAVSVGELEHODWRO
VAVAVLADGEDTYLALGRVSVNHEAVEILRDAGLHPHGLESHDVAQAAYVD
AVAGAVAVLRRRAARLSDVGRSPBEIATQLLVPTDPAVSVLRLDAQAQERG
ACBAARYLERVEAEKDPDVLRLGKLAETDPARSVTLHEASLITDVARAT
AVOGLICLAAVQSPDGRALTEALDALDELGPEDRRELRITLTESALLVSGDE
KATLPILRTESGLTPQPGDTPAQROQLMSTLSAEGGAEITGKARARARAGV
PLGWSLPLPTSLALSDENEAEEVLETVLRSGSDTAAYTVIALSTRSLRENG
AVPDAMADQTALEILGOERMGDTATMPHTAVASVLTERRDPRALDALIKRPHLD
RFVWEYHWYLMARGLAADGDGLALOVFGSCGSAEAGLTNPVLA.PMWLETCIL
GEARGEEBGRAAAHGRLAERWGTARLGYALARGAARGTAPGALREAVALLAD
SPARSHGRLILLGRALVEDGAVRSGREHLRAVGLARCCGVALARARDLILAG
GMRREVTASPLDMLTGERTVAGLVASGAENREVAESLFTVATVELHLITSYRKLV
ARRGDLTEALREMGATARPAAERRPGHAKRNP"
/3522..14898
/gene="fscM1"
/3522..14898
/gene="fscM1"
/note="glycosyltransferase responsible for mycosamine
attachment"
/codon_start=1
/translate=1
/product="fscM1"
/protein_id="AA082555.1"
/db_xref="GI:34766442"
/translation="MDSARPIFLVSLPESGLNPLVLVAGELARGVDLWFATDEH
RREVEVALDSVKSVEFASLGEVPELSAVWDEVREYQOSRFAHAAVYQOSR
ALQARKYRELEQVDEVPRALWVDCVAGVDLARGIPYLANTPFAVASNTLSTN
PFGASTPKSPVNSGLPARMSVROKLANTLTKMTLTCMFLHPDMAALLRBAARL
ELGIAPNMATRVDEAAAYVCSSVAELDFPDI.PDVSVLGAFLPLPPLAPDDDEVTR
WLDQSSVYMGFTLITRLREVAALVEVARRMSGTHQFLMWLTPXEQOHLPEASOL
PDLRVESWVPSQDLVLAHPNVSPFSSHGGNAHGVYFGKQVVRPLMWVCFDQAV
RGHDFGISLTLDKPHTVDDVDVDKLTVSDBAFTTEERLGLARGAAGRAAADL
VTGLPAA"
/14953..16011
```

```
CDS
/gene="fscM1"
/14953..16011
/gene="fscM1"
/note="putative aminotransferase responsible for
mycosamine biosynthesis"
/codon_start=1
/translate=1
/product="fscM1"
/protein_id="AA082556.1"
/db_xref="GI:34766443"
/translation="MAFTPVSRPALDRELEVSDAVSGWISSQGPVYRREFAFA
ENMGVAGVACSSTGATLTRLALNIGDEVTYVEFTMVASMAVTVGATPVFD
CGDDNIDVTRIEEKTTPRTAVRAPHVGRCDMDAMVLAQVNLRVEDSABHG
VRVVGILACFSLPANKITTAGGVGLTDDPLRALATIRAMA.PTRDSFLHXILAY
NRTMTMQAVNALAQTERLDELATRRREIARIDAGLKLPGITLMPADVLMYDIL
AERRELRHLDRIGLETLPFKPMSRQGPIDVPWPTLNAHRSBDGLYLPHTGLT
AADQEVITGAVDPFYRAG"

Alignment Scores:
Pred. No.: 6,52e-76 Length: 138203
Score: 1766.00 Matches: 381
Percent Similarity: 63.93% Conservative: 71
Best Local Similarity: 53.89% Mismatches: 209
Query Match: 49.43% Indels: 46
DB: 1 Gaps: 13

US-10-089-514-2 (1-686) x AY310323 (1-138203)
QY 1 MetArgThrLeuLeuIleAspAsnYrAspSerPheThrHisAsnLeuPheGlnIyrIle 20
18610 ATCGGACCCCTTCTGTCGACAACTACGACTGTTACCTACCACTCTTCCACTACCTC 18669
QY 21 GlyAlaIleThrGlyGlnProProValValProAsn--AspAlaAspTrpSerArg 39
18670 TCCCGGCGCAACGCGCGGACCGGAGTCACTCCGACGACGACCGCGCTCCGCGCC 18729
QY 40 LeuProValGluAspPheAspAlaIleValValSerProGlyProGlySerProAspArg 59
18730 GGTCTGCTGACCGCTTCCACAACTGCTCTCCCGGCGCGGCTACCCACACCGC 18789
QY 60 GluArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValIleGly 79
18790 CCGCGCGACTTCCGCTGCTGCGCGCGGACGCGGAGAGGCGCGCTCCGCTCGCG 18849
QY 80 ValCysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaPro 99
18850 GTCTGCTGCGCGCAACGAGGCACTGCGCTCCGACGCGCGCGGCTGCGCGCGCC 18909
QY 100 GluProMetHisGlyValArgValSerGluValValGHisThrGlyGluAspValPheArgGly 119
18910 GAGCCCGCTCACGCGCGCACTTCCGCGGTACGCGACGAGCGGACCGGCTCTTCCAGGG 18969
QY 120 LeuProSerProPheThrAlaValArgYrHisSerLeuAlaAlaThrAspLeuProAsp 139
18970 CTGCGCGACCGCTGAGAGTGTGCGTACCTCTCTCGGTGACGGAACCTGCGCGCG 19029
QY 140 GluLeuGluProLeuAlaTrpSerAspAspGlyValValMetGlyLeuArgHisArgGlu 159
19030 GAGCTGAGCGCACCGCTGCTGCGAGAGCGGGGTCTGATGCGCTGCGCACCGCACG 19089
QY 160 LysProLeuTrpGlyValAlaGlnPheHisProGluSerIleGlySerAspPheGlyArgGlu 179
19090 CTGCGCGCTTGGGGGTGCGAGTTCCACCCGAGTGCATCGGACCCGAGCGCACCGG 19149
QY 180 IleMetAlaAsnPheArgAspLeuAlaLeuAlaHisHisArgAlaArg----- 195
19150 CTGCTGCGCACTTCCGCACTTCCACCGAGCGCCACGCGCCGACCGCGCGCGCGG 19209
QY 196 ---ArgHisGly-----AlaAsp 200
19210 GCGGGGACGCGACGCTCCCGCGCGCGCGCGCGCGGACGACGCGCGCACCGCGCG 19269
QY 201 SerProTyrGluLeuHisValArgArgValAspValLeuPro-----AspAlaGlu 217
```

gene



```
Db      19270 ACCCCACGCGGCTCCGGTCC---ATCGCCGAGTCGCTCCCAACGCGCTGGGACGCGCGAG 19326
Qy      218  GLuVal-----ArgarglyCyLeuProglYgluIyThrThrPheTPr 232
Db      19327 GTCCGCTTCGACTCGCTGTTCCGACCGGCGACCC-----TTCTGG 19371
Qy      233  LeuAspSerSerSerValLeuIglYAlaSerArgPheSerPheLeuIglYAspAparXg 252
Db      19372 CTGACAGAGCGCGCTCCCGGGGGGAGCTGGGCGAGTCTCCGTATGGCGCGACTCA 19431
Qy      253  GLYProLeuAlaIgluYrleuThrTYrArgValAlaAspGLYValSerValArgGLY 272
Db      19432 GGTCCCTCGCCCGGACCGGCAAGCGCGACGTCAGCGCGGAACGGTCAACGCTGAAGA--- 19488
Qy      273  SerAspGLYThrThrThrArgThrArgArgProPhePheAsnTYrleuIgluIgluInleu 292
Db      19489 GCCGACGGCGCCGACGACGCGTCGAGAGCGCCTTCCTGACCTGGCTGGAACGACCTG 19548
Qy      293  GLuArgArgArgValProValAlaProGLuLeuProPheIgluPheAsnLeuGLYTYrVal 312
Db      19549 GCGGGGCTGGCGACCGAGG---CCGAACTTCCTTCGCGCTTCGCGCTCGGCTGGTGC 19605
Qy      313  GLYTYrleuGLYTYrGLuLeuYsAlaIgluThrThrGLYAspProAlaHisArgSerPro 332
Db      19606 GGTGCTGGGCTAGCAGGTTGAAGCCGAGTCGACGCGGACGCGCGGACCGCTCACCC 19665
Qy      333  HisProAspAlaAlaPheLeuPheAlaAspArgAlaAlaLeuAspHisGLuIgluY 352
Db      19666 GATCCGACGCGCTGCTGCTTCGCGACCGGGGCTGTGTGTGACCAACCCACCCCG 19725
Qy      353  CysCysTYrleuLeuAlaLeuAspArgArgLYHisAspAspGLYAlaArgAlaTrPLeu 372
Db      19726 ACCACTACTCTGTTGGCGCTGTGGAGAGACGACGCGGACCGAGCGACTGGCTTC 19785
Qy      373  ArgGLuThrAlaIgluThrleuThrGLYleuAlaValArgAlaProAlaIgluProThrPro 392
Db      19786 GCGGCGGCTCCGCGCACCTCGAGGCGATCGCGGGCG-----GAGCCCGAGCGG 19836
Qy      393  AlaMetValPheGLYleuProGLuAla---AlaAlaIgluPheGLYProleuAlaArgAla 411
Db      19837 TGC-----CCGAGGCGCGCGTGTGACGACGCGGTCCGG---GAGCTG 19878
Qy      412  ArgHisAspLysAspAlaTYrleuLYsArgLYAspGLuCyLeuYsGLuIleArgAsn 431
Db      19879 CGCCACGACCGGAGACGGCTACCTGAAGCTATGACGCTCGCCAGAGGAGATAGCCGCC 19938
Qy      432  GLYLeuSerTYrGLuIleCyLeuThrAsnMetValThrAlaProThrGLuAlaThrAla 451
Db      19939 GGGGAGACCTACGAGGCTGTGCTGACCAACATGCGAGGCGGACACCGACCTCACCCCG 19998
Qy      452  LeuProleuTYrSerAlaLeuArgAlaIleSerProValProTYrGLYAlaLeuLeuGLu 471
Db      19999 TGGGCGGCTACCGCGGCTGCGCGCGGCTGAGCCCGCGCTTCGCGCTTCGCTG 20058
Qy      472  PheProGLuLeuSerValLeuSerAlaSerProGLuArgPheLeuThrIleGLYAlaAsp 491
Db      20059 TTGGGCCCATGGCGGCTGAGAGCTCTCCGAGCGGTTCTCTGGCATGACCGGAC 20118
Qy      492  GLYGLYValGLuSerLYsProIleLYsGLYThrArgProArgGLYglYThrAlaIgluGLu 511
Db      20119 GGGCGGATGAGTCGACCAAGCGGATCAAGGAGGACCGCGGCGACCGCGCGACGAGAG 20178
Qy      512  AspGLuArgLeuArgAlaAspLeuAlaGLYArgGLuLYsAspArgAlaIgluAsnLeuMet 531
Db      20179 GACGCGCGCTGCTACGCGCGCTGGCCACTGCGAGAGAGACCGCGCGGAACTGATG 20238
Qy      532  IleValAspLeuValArgAsnAspLeuAsnSerValCysAlaIleGLYSerValHisVal 551
Db      20239 ATGCTGACCTGTGTCCGACGACCTGGGGCGGTGCGCCGAGAGTCCGCTGCTGCTGCC 20298
Qy      552  ProArgLeuPheGLuValIgluThrTYrAlaProValHisGLuIleuValSerThrIleArg 571
```

```
Db      20299 GACCCGGTTCACGCTCGAGACGTAACGCGACCTGACCAACGCTGCTAGTACCTCACG 20358
Qy      572  GLYArgLeuArgProGLYThrSerThrAlaAlaCysValArgAlaAlaPheProGLYGLY 591
Db      20359 GCGGCGCTCGGAGAGACGACGCCGCGTGGCGGCTCCGGGCGCTTCCCCGGCGGA 20418
Qy      592  SerMetThrGLYAlaProLYsLYsArgThrMetGLuIleIleAspArgLeuIgluIglY 611
Db      20419 TCGATGACCGGGGCGCGGAGATCCGACCATGACGATCATCGACCGGCTGGAAAGCGGG 20478
Qy      612  ProArgGLYValTYrSerGLYAlaLeuGLYTYrPheAlaLeuSerGLYAlaAspLeu 631
Db      20479 CCGCGGAGTGTACTCCGGCGCCCATCGGCTACTTCTCCCTCACCGCGCGGTGACCTG 20538
Qy      632  SerIleValIleArgThrIleValLeuAlaAspGLYGLuAlaIgluPheGLYValIglYGLY 651
Db      20539 TCCATGTATCCGACCGGTGTGCTTCACGGCGGCGGACGCTGCTGACGCGCTGGCGGC 20598
Qy      652  AlaIleValSerLeuSerAspIgluIgluIgluIgluThrGLuThrValLYsAlaArg 671
Db      20599 GCGGTATGCGCCTCTCCGACCGGCGGCGACGAGTTGAGAGAGACGCGGCTCAAGCGCGC 20658
Qy      672  AlaMetValThrAlaLeuAsp 678
Db      20659 CCGCTGCTGCTCCTCCGAC 20679

RESULT 11
AP006618_17/c
WPCOMMENT
Sequence split into 61 fragments LOCUS AP006618 Accession AP006618
Fragment Name      Begin      End
AP006618_00      1      110000
AP006618_01      100001    210000
AP006618_02      200001    310000
AP006618_03      300001    410000
AP006618_04      400001    510000
AP006618_05      500001    610000
AP006618_06      600001    710000
AP006618_07      700001    810000
AP006618_08      800001    910000
AP006618_09      900001   1010000
AP006618_10     1100001   1210000
AP006618_11     1200001   1310000
AP006618_12     1300001   1410000
AP006618_13     1400001   1510000
AP006618_14     1500001   1610000
AP006618_15     1600001   1710000
AP006618_16     1700001   1810000
AP006618_17     1800001   1910000
AP006618_18     1900001   2010000
AP006618_19     2000001   2110000
AP006618_20     2100001   2210000
AP006618_21     2200001   2310000
AP006618_22     2300001   2410000
AP006618_23     2400001   2510000
AP006618_24     2500001   2610000
AP006618_25     2600001   2710000
AP006618_26     2700001   2810000
AP006618_27     2800001   2910000
AP006618_28     2900001   3010000
AP006618_29     3000001   3110000
AP006618_30     3100001   3210000
AP006618_31     3200001   3310000
AP006618_32     3300001   3410000
AP006618_33     3400001   3510000
AP006618_34     3500001   3610000
AP006618_35     3600001   3710000
AP006618_36     3700001   3810000
AP006618_37     3800001   3910000
AP006618_38     3900001   4010000
AP006618_39     4000001   4110000
AP006618_40     4100001   4210000
AP006618_41
```



AP00618\_42 420001 431000  
AP00618\_43 430001 441000  
AP00618\_44 440001 451000  
AP00618\_45 450001 461000  
AP00618\_46 460001 471000  
AP00618\_47 470001 481000  
AP00618\_48 480001 491000  
AP00618\_49 490001 501000  
AP00618\_50 500001 511000  
AP00618\_51 510001 521000  
AP00618\_52 520001 531000  
AP00618\_53 530001 541000  
AP00618\_54 540001 551000  
AP00618\_55 550001 561000  
AP00618\_56 560001 571000  
AP00618\_57 570001 581000  
AP00618\_58 580001 591000  
AP00618\_59 590001 601000  
AP00618\_60 600001 602125  
Continuation 118 of 61 of AP00618 from base 170001 (AP00618 Nocardia farcinica IFM 1)

## Alignment Scores:

Pred. No.:	1,25e-69	Length:	110000
Score:	1635.50	Matches:	364
Percent Similarity:	60.40%	Conservative:	60
Best Local Similarity:	51.85%	Mismatches:	241
Query Match:	45.77%	Indels:	37
DB:	1	Gaps:	13

US-10-089-514-2 (1-686) x AP00618\_17 (1-110000)

QY 1 MetArgThrIeuLeuLysAspSerPheThrHisLeuPheGlnTyrIle 20  
DB 54159 ATGCGACGCTGTGTGACGACACTACGACTCTTCACTCAACCTGACAGCTGATC 54100  
QY 21 G1G1uAlaThrGlnPheProValValValProAspSer---AlaAspTProSerArg 39  
DB 54099 AGTGAAGTGAACGGGGTGCAGACCGACCTGCTGCGCAACAGACCGACCGCGCG 54040  
QY 40 LeuProValG1uAspPheAspAlaIleValValSerProG1yProG1ySerProAspArg 59  
DB 54039 CTGACCTCGCCCGCTGACGACCAATCGTCTCCCGCGCGCGCGCGCGCGCGCG 53980  
QY 60 G1uArgAspPheG1yLysSerArgAlaIleThrAspSerG1yLeuProValLeuG1y 79  
DB 53979 ACCCGGACGTCGGCTGTCGCGACGCGTACGACCGCGCGCTGCCAAGCTGCTCGC 53920  
QY 80 ValCysLeuG1yHisGlnG1yIleAlaGlnLeuPheG1yG1yThrValG1yLeuAlaPro 99  
DB 53919 GTGTGCTGGGGACACGAGGATCGTGTGCGCGCGCGCGCGCGCGCGCGCGCG 53860  
QY 100 G1uProMetHisG1yArgValSerG1uValArgHisThrG1yG1uAspValPheArgG1y 119  
DB 53859 GTACCCCGGACGCTCTCGACCGGGTGGCGACGCGCGCGCGCGCGCGCGCG 53800  
QY 120 LeuProSerProPheThrAlaValArgTyrHisSerLeuAlaIleThrAsp---LeuPro 138  
DB 53799 CTGCCGACGATTCACGCTGCGCGTACCTCGCTGCGCGCGCGCGCGCGCGCG 53740  
QY 139 AspG1uLeuG1uProLeuAlaTProSerAspAspG1yValValMetG1yLeuArgHisArg 158  
DB 53739 CCGCGCTCGAGGTACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 53680  
QY 159 G1uLysProLeuTProG1yValGlnPheHisProG1uSerIleG1ySerAspPheG1yArg 178  
DB 53679 CAGCGCGCGACGTGGGTGTGAGTTCATCCGAGTCGGTCCGCGCGCGCGCGCG 53620  
QY 179 G1uIleMetAlaAspPheArgPheLeuAlaLeuAlaHisHisArg-----193  
DB 53619 GCGCTCTGCGCAACTTCGCGCACTGACCTGCGCGCGCGCGCGCGCGCGCG 53560  
QY 194 -----AlaArgArgHisG1yAlaAspSerPro-----202

DB 53559 CGTCCGCGGACACCGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 53500  
QY 203 -----TyrG1uLeuHisValArgArgValAspValLeuProAspAlaG1uValArg 220  
DB 53499 CTGCGGACGAGGTGATCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 53446  
QY 221 ---ArgG1yCysLeuProG1yG1uG1yThrThrPheThrLeuAspSerSerValLeu 239  
DB 53445 CTCTACGCGCGCTGCGCG-----ACGCGCTTGGCTCGACGCGCAACGCTGACG 53395  
QY 240 G1uG1yAlaSerArgPheSerPheLeuG1yAspAspArgG1yProLeuAlaG1uTyrLeu 259  
DB 53394 CCGGCGTGGACCGGTTCTCTCTTCTCGGTGACGCGAGCGACCGCTGGCGAGGTG 53335  
QY 260 ThrTyrArgValAlaAspG1yValValSerValArgG1ySerAspG1yThrThrArg 279  
DB 53334 CGTTACCGGCTCGGTGCGCGACGTGCGCGGTGAGAGCGCGCGGAAACGCGAGCG--- 53278  
QY 280 ThrArgArgProPhePheAspThrIleLeuG1uG1uG1uLeuArgArgValProVal 289  
DB 53277 GTCCCGCGGACGCTGCTGGGTATTCGTCCACCGACTCGCGAGCGACCTGAGCTG 53218  
QY 300 AlaProG1uLeuProPheG1uPheAspLeuG1yTyrValG1yTyrLeuG1yTyrG1uLeu 319  
DB 53217 ---CCGCGACCTGCGCTTGCATCTTCGCGCGGATTAAGTGTGCTACCTCGGCTACGAGCTC 53161  
QY 320 LysAlaG1uThrThrG1yAspProAlaHisArgSerProHisProAspAlaAlaPheLeu 339  
DB 53160 AAGCGCGACTGCG 53101  
QY 340 PheAlaAspArgAlaIleAlaLeuAspHisGlnG1uG1yCysSerTyrLeuLeuAlaLeu 359  
DB 53100 TTGCGCGACCGGTGTATCTGTGTCGACAGTCGCGCGGCAACCCATCTGTGTGGCTC 53041  
QY 360 AspArgArgG1yHisAspAspG1yAlaArgAlaTProLeuArgG1uThrAlaG1uThrLeu 379  
DB 53040 -----ACGAGACGACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 52987  
QY 380 ThrG1yLeuAlaValArgAlaProAlaG1uProThrProAlaMetValPheG1yIlePro 399  
DB 52986 GAAACGCTG-----CCCGGTGGCGAATCCGCGGAAATTG---GCGATCGAG 52942  
QY 400 G1uAlaAlaAlaG1yPheG1yProLeuAlaIleArgAlaArgHisAspLysAspAlaTyrLeu 419  
DB 52941 TCCGCGCGCGCTGCGTGGCGCGCG---CTGAACCGGCGCGCGACCGCTACCTC 52888  
QY 420 LysArgIleAspG1uCysLeuLysG1uIleArgAsnG1yG1uSerTyrG1uIleCysLeu 439  
DB 52887 GCGGACATCGCGCGCTGCTTCAAGTATCTGCGCGCGGAGATCTTCAAGATCTGCTG 52828  
QY 440 ThrAspMetValThrAlaProThrG1uAlaThrAlaLeuProLeuTyrSerAlaLeuArg 459  
DB 52827 ACCGACAGCTACCGTGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 52768  
QY 460 AlaIleSerProValProTyrG1yAlaLeuLeuG1uPheProG1uLeuSerValLeuSer 479  
DB 52767 CCGTCGAATCCGCGCGCTTATCCGCTTACCTGCGCTTGAACGATCCAGATCGCTCC 52708  
QY 480 AlaSerProG1uArgPheLeuThrIleG1yAlaAspG1yG1yValG1uSerTyrProIle 499  
DB 52707 TCTTCAAGGACGCTTCTTCAAGTACGACGCGCGCGCGCGCGCGCGCGCGCGCG 52648  
QY 500 LysG1yThrArgProArgG1yG1yThrAlaG1uG1uAspG1yArgIleAspLeu 519  
DB 52647 AAGGCGACCG 52588  
QY 520 AlaG1yArgG1uLysAspArgAlaG1uLeuMetIleValAspLeuValArgAspArg 539  
DB 52587 GCGGACAGTCCCAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 52528  
QY 540 LeuAspSerValCysAlaIleG1ySerValHisValProArgLeuPheG1uValG1uThr 559  
DB 52527 CTGCGCGGAGTTCGACAGTTCGCGAGCTGACGCTGACCTGATGCGCGCGAGTGC 52468



OY		560	TyrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuAhrProGlyThrSer	579
Db		52467	TATTCACCGTCCACCAGCTGGTGTCCAAGTGAGCGGCACCTTGCGCCGGAAGTGCAC	52408
OY		580	ThrlaalaCySValaArgAlaaIaphPheProGlyVGISerMetThrGlyAarPolySuLyS	599
Db		52407	GCCATCAGACTGCCGTGGCGGCTGTTCCCGGCGCTCATAGACGGCGCGGAAGCTG	52348
OY		600	ArgThrMetGluIlelleAPaArgLeuGluglyProArgGlyValIyrsSerGlyAla	619
Db		52347	CGCACGATGTGAGATCAATCGACGAGCTGGAAGAGCGCGCGCGGGGTCTAACTCCGACCC	52288
OY		620	LeuGIyrrPheAlaIaleuSerglyValaIaAspLeuSerlleValIleArghrIleVal	639
Db		52287	ATTCGGCTTCCTCGGACTCGGGGACCGCGGACCTCAACTGTCTATCCGACCGGGGTG	52228
OY		640	LeuAlaaspGlyGlnAlaIagluPheGlyValIGlyVGlyValallelserLeuSerAapGln	659
Db		52227	CGGACACACGGGGCGCTGGCGGATGGGGGGGGGGCGCGCATCTGTGGATTCCATCCG	52168
OY		660	GIuGIuPheThrTrgIuThrValValIyAlaArgAlaMetValThrAlaLeuaspGly	679
Db		52167	GCGAGGAGGTATCCGACGATGTGTGGAGAGCGCGCGGACCCAGCGCGGCGCACCGCC	52108
OY		680	SeralA	681
Db		52107	GGGGCC	52102
	RESULT 12			
	LOCUS	AP003593	333550 bp DNA linear BCT 28-NOV-2001	
	DEFINITION	Nostoc sp. PCC 7120 DNA, complete genome, section 13/19.		
	ACCESSION	AP003593 BA000019		
	VERSION	AP003593.1 GI:17132540		
	KEYWORDS	.		
	SOURCE	Nostoc sp. PCC 7120		
	ORGANISM	Nostoc sp. PCC 7120		
	REFERENCE	Bacteria; Cyanobacterieia; Nostocales; Nostocaceae; Nostoc.		
	AUTHORS	1 Kaneko,T., Nakamura,Y., WolK,C.P., Kuritz,T., Sasamoto,S., Watanabe,A., Iriguchi,M., Ishikawa,A., Kawashima,K., Kimura,T., Kishida,Y., Kohara,M., Matsumoto,M., Matsuno,A., Muraki,A., Nakazaki,N., Shimpo,S., Sugimoto,M., Takazawa,M., Yamada,M., Yasuda,M. and Tabata,S. Complete genomic sequence of the filamentous nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120 DNA Res. 8 (5), 205-213 (2001)		
	JOURNAL	MEDLINE		
	PUBMED	11759840		
	REFERENCE	2 (bases 1 to 333550)		
	AUTHORS	Kaneko,T.		
	TITLE	Direct Submission		
	JOURNAL	Submitted (02-May-2001) Takakazu Kaneko, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail:kaneko@kazusa.or.jp, URL:http://www.kazusa.or.jp/cyanobase/ Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934) Location/Qualifiers 1..333550 /organism="Nostoc sp. PCC 7120" /mol_type="genomic DNA" /db_xref="taxon:103690" /note="synonym:Anabaena sp. PCC7120" complement(52..1455) /gene="ai13408" complement(52..1455) /gene="ai13408" complement(52..1455) /note="ORF_ID:ai13408" /codon_start=1 /transl_table=11 /product="nitotinamide nucleotide transhydrogenase,		
	FEATURES			
	source			

subunit beta"  
/protein\_id="BAB75107.1"  
/db\_xref="GI:17132541"  
/translation="MSDFLPTGIQITLYVAASLFIQLKKLGSPATARNGVIGAVGM  
LLAIATVAITLDHVNEMWILLGLAIGSGIGAIAYKVQMTPEPMQVGLNGLGGAAS  
LVAAPFMRLLIDAGAPVLDVNIISMLDVLIGVTFPGSFLAFKLGILLSGSPITFP  
LQAPNNILLGGVAGVYLIVTPDSLPVLIYAVASVLVGMFVIPIGGDMPVIS  
ILNLSISGIAAAAGFVVMNNMLITAGLVASGLILTEINCKAMNSLSVLSARG  
VSAGAGVATITQTVHASIDPEEGAMLLGARVSIVIPQIGMAVQAQSVBELDQ  
LERMVDVATIHVPVAGMPEGHNNVLLAEANVPYQIVYDMDINPQEQADVALVIGA  
NDVAVPARSSTFASPIGMPILLEVDRAKHTIVIKRGSAGFAGVDNELFYKDKTLMLF  
GSAKVMVAKLVSEVKOL"  
complement(1452..1745)  
/gene="as13409"  
complement(1452..1745)  
/gene="as13409"  
/note="ORF\_ID:as13409"  
/codon\_start=1  
/transl\_table=1  
/product="nicotinamide nucleotide transhydrogenase,  
subunit alpha"  
/protein\_id="BAB75108.1"  
/db\_xref="GI:17132542"  
/translation="MTTEILALFVPLASFIGEVINKIPITLHPILMSGNASISGI  
AVIGAILASGRNTNLSYIIGLIYATVATVAVAGFLVTBMLQPFKKREVKA"  
complement(1785..3020)  
/gene="a113410"  
complement(1785..3020)  
/gene="a113410"  
/note="ORF\_ID:a113410"  
/codon\_start=1  
/transl\_table=1  
/product="nicotinamide nucleotide transhydrogenase,  
subunit alpha"  
/protein\_id="BAB75109.1"  
/db\_xref="GI:17132543"  
/translation="MTPLISKVADPAFIRIYSSDFMRIRAVAKIEVCERRVALNPPT  
VARLQKQLEWSEVAGAEERSFPDDAAVVAAGVTVGDSQKLWETDVLIVKSPQDR  
EDGRLEIDLEBEGVILSFNLPLDNPVAAQVQVNRKVTALSMELIPTTQAQSDRLAS  
SOASLAGKAVKILAAALPKFYPMILTAAGTIAPAKVIFINGAGVAGIQAATARALCA  
VVEADIDPAAKEVOSIGAKFEVTLKEETTAGVAKGEISESKACTQDEVVAEHVK  
NSDVITTAQVGRKAPVLIYAEVAVQMKGSVTVDLAEOGNGCACTDGGKQIVNNG  
VTIITPIVBSMRYHVASQLYSKULSTLLQILLKDKALQVDFDDITIDACITIHAGSI  
RNQRKDALQASTQLVARD"  
3627..4199  
/gene="a1r3411"  
3627..4199  
/gene="a1r3411"  
/note="ORF\_ID:a1r3411  
hypothetical protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAB75110.1"  
/db\_xref="GI:17132544"  
/translation="MLIVNLVITIKIPMRRLTMRLLSGLAVTSVVIAGLPITITS  
QSLPEPFLSSVSKENMLPRLDFGQGSNSWDYILRIPAQKKMLAAQPAITVPHY  
KGSFETKRIEIRVGDKKVPSSEVMDREGRIELIFPEEPVAGSRVELVLSNVNPAF  
GCTEFYNCQIISGVPVPLRLRYLGTWILIN"  
4346..4480  
/gene="rpi134"  
4346..4480  
/gene="rpi134"  
/note="ORF\_ID:as13412"  
/codon\_start=1  
/transl\_table=1  
/product="50S ribosomal protein L34"  
/protein\_id="BAB75111.1"  
/db\_xref="GI:17132545"  
/translation="MQRTLGSTNRKRRKSTSGFRARMPTPDGRNVIRARRKGRHRLSV  
"  
4506..4928  
/gene="a1r3413"  
4506..4928  
/gene="a1r3413"  
4506..4928  
/gene="a1r3413"



```
/gene="a1r3413"
/notes="ORF_ID:a1r3413"
/codon_start=1
/translate=1
/product="ribonuclease P"
/protein_id="BAB75112.1"
/db_xref="GI:17132546"
/translation="MALPKANRLKSRHDPQAVPREGLRNRSHPFLRALKPSARKSS
LDTAAQDPVDEVONI PSTLIGVISITKSKRAVVRNRI GRQITAAQQLPRLAIGM
KLVIVKPTAKSKSCSQDFLQELBOLAEVLVHGS"
4915..5298
/gene="a1r3414"
4915..5298
/notes="ORF_ID:a1r3414"
/hypothetical protein"
/codon_start=1
/translate=1
/protein_id="BAB75113.1"
/db_xref="GI:17132547"
/translation="MGIRESVYEGGPHIGDLINILVGLTVVGIPLAVGATYRALML
RRIITRRISVMGMMGRDSVITYSEITTKIVKVPGLGIWGMVVTLNKGSLLELA
VENFREVDYDINKIAKNPQYAK"
5513..6655
/gene="a1r3415"
5513..6655
/notes="ORF_ID:a1r3415"
/hypothetical protein"
/codon_start=1
/translate=1
/protein_id="BAB75114.1"
/db_xref="GI:17132548"
/translation="MDPGLGFLSNMVLPIIDFFYVDSYGLAIVALLIIFALYLP
LSGQIRMKMKRIYQPLQKMAEI KERYKDPQKQEBMNAVQKEFGNPLAGCLPL
LLOMPVLLALFATLRSPFAGVSVNLQIFPSEQIERLPQAFAPRPQIVYADGEL
TKVTALLPGNKLAIVGSEKTIQYOLEGKFPVLLAEHPKSKSPMKITKEDRKI
DAENIEALQPGDVTIGCTPGIADKGPFLDALGRVADIDPNTIINDYMIYFP
GISLYSOMLSGNSSGGNPDQDTNKKITPVIISGNFLPPLPAGVLMIVIGNVQT
LOTYILSRPLPEIKIVETOKQAVAEOKTLPSPKSKKKTG"
6659..7177
/gene="a1r3416"
6659..7177
/notes="ORF_ID:a1r3416"
/hypothetical protein"
/codon_start=1
/translate=1
/protein_id="BAB75115.1"
/db_xref="GI:17132549"
/translation="KSNNPWQRCQQLQSLLELTGVSAEIHGSLFETQASHNDSQEID
GWLITIDETNLNFOQIOTLIGADGSLVDAI QYLVNSTLININQEBQSAIYVLENGYR
VKQAEIQKIAETFAEAOVRSVGOEIEIKSLSAERLVAHTFLKDFGLTFPSRKEPH
RHLVVRPAVAMEL"
7339..7848
/gene="a1r3417"
7339..7848
/notes="ORF_ID:a1r3417"
/hypothetical protein"
/codon_start=1
/translate=1
/protein_id="BAB75116.1"
/db_xref="GI:17132550"
/translation="MDAIFIPQLTKAPERTEEVQKFLPGLFETTLTPVGLVRIHQG
NLYQVAAQBAIITCTCRCLQYORLAVNTQEVMLDEBAVEQNLPLERVAEAD
LLETVPNGYFVSEMLYEQMCLALPQRQCDINCIGILINDSGESSBGGINDRMGCL
KALKQLRPC"
complement(7972..9447)
/gene="a113418"
complement(7972..9447)
/gene="a113418"
/notes="ORF_ID:a113418"
```

```
hypothetical protein"
/codon_start=1
/translate=1
/protein_id="BAB75117.1"
/db_xref="GI:17132551"
/translation="MTFMPFLLLIGIATVLMVORSVNIITKPTIMLMLVMTPLVILT
GMTLVYGVKQPPPALIFWPSIICLLVYLLIOMWGLRDPDTAQTOTESGQANPT
AEVPRVRIPEPTBEQOLRCFPMSTIYYONIEYRPAVLCROLRTPPOAXOIRAN
TEQOFDRFLLIPOSEFNKRPFFVLPNSOAKANKRSSEQLTRPGLALLIVATLYT
TVLVGAKIAGIDPTRLQSDPKLFGSLPYALMLTLGHEGHWYLVARFKIRSTLP
YFPIPFLEGTGAFIQMRSPINRKALPDVGIAGLAFIATLPVIVGLAKSDLPV
LTENTSILNPDALNPXYSIIVALLAKLIGSALTAKLAIIDLHPVAVAGFLIVTALN

Alignment Scores:
Pred. No.: 1,176-67 Length: 333550
Score: 1606.00 Matches: 335
Percent Similarity: 62.204 Conservative: 101
Best Local Similarity: 47.794 Mismatches: 227
Query Match: 44.954 Indels: 38
DB: 1 Gaps: 9

US-10-089-514-2 (1-686) x AP003593 (1-333550)
QY 1 MetArgThrIleuLeuIleAspAsnTYrAspSerPheThrHisAsnLeuPheGlnTYrIle 20
      ::::::::::::::::::::
DB 33355 ATCAAAACATTATTTATTCACACTGCACTCTTATCTTTATCTTACACGATGATT 33414
QY 21 GlyIuAlaThrGlyGlnProValValAlaProAsnAsp---AlaAspTPSerArg 39
      ::::::::::::::::::::
DB 33415 GCTGAAGTCACAGAGAGTGGCCCTTGTATTCATATGACAAATTAAGATTGTATGAA 33474
QY 40 LeuProValGluAspPheAspAlaIleValIleSerProGlyProGlySerProAspArg 59
      ::::::::::::::::::::
DB 33475 TTAACAACTGTTTATATATATATATGATTCATTCCTCGGGCGTGTCTGCAAAAC 33534
QY 60 GluArgAspPheGlyIleSerArgArgAlaIleThrAsp---SerGlyLeuProValLeu 78
      ::::::::::::::::::::
DB 33535 CCTAAAGATTGCGATATTTGTCACAGATACATCAAACTATTATTAATGATGTTCCCTTGG 33594
QY 79 GlyValCysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThyValGlyLeuAla 98
      ::::::::::::::::::::
DB 33595 GGTGTTGCTTGTCGTCACAGAGATGTCATTATATATGATGTCGTCGTCGTCGTCGTCGTC 33654
QY 99 ProGluProMetHisGlyArgValSerGlyValArgHisThrGlyGluAspValPheArg 118
      ::::::::::::::::::::
DB 33655 CCAGAACCTTAACATGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 33714
QY 119 GlyLeuProSerProPheThrAlaValArgTYrHisSer---LeuAlaIleThrAspLeu 137
      ::::::::::::::::::::
DB 33715 GAAATTCCTTACACCTTTTCAAGTTGTCGCTATTCATTTCTTGCTGCTGCTGCGAAGATTGG 33774
QY 138 ProAspGluLeuGlnProLeuAlaIleTyrSerAspAspGlyValValMetGlyLeuArgHis 157
      ::::::::::::::::::::
DB 33775 CCAGTGTGTTTGAAGAAAGTACGCTGGAAGAACGGGTGTGTAATGGGGTGGCTCAT 33834
QY 158 ArgGluLysProLeuTYrGlyValGlnPheHisProGlySerIleGlySerAspPheGly 177
      ::::::::::::::::::::
DB 33835 CGTCATTTCCTCATTTGGGGGGGGGCGAGTTTCCAGAACCTTATCTGATGAGTATGGA 33894
QY 178 ArgGluIleMetAlaAsnPheArgSerLysAlaLeuAlaHisHisArgValArgArgHis 197
      ::::::::::::::::::::
DB 33895 TATTAATTCCTGAAAGAACTTCCAA-----GCCATTACGAAAGATTATCCGCGCAAAAT 33948
QY 198 GlyAlaAspSerPro----- 202
      ::::::::::::::::::::
DB 33949 GGTAGACATCGCGCCCAATTAAGAAAGCAAAAGTTTCTATTTCTCCCTAATCAGCTATT 34008
QY 203 -----TyrGluLeuHisValAlaArgValAlaAspValLeuProAspAla 216
      ::::::::::::::::::::
DB 34009 CCAGCGCCATCCCAAGAAATTTACTGTTGTCACCAAAACCTGATTTATACCTTGATACA 34068
QY 217 GluGluValAlaArgArgGlyCysLeuProGlyGlyGlyIleThrPheThrLeuAspSerSer 236
      ::::::::::::::::::::
```



Db 34069 GAACAGACATTTGTCATCTTTTGTGTAAGATCCATCGCTTTTGGCTTGATGATGAC 34128  
 QY 237 SerValIleuGlulGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAla 256  
 Db 34129 CTTGTGACACCAAGATTTATCCCGTTTCTTTTATGAGGAGGAGCGGGGATTAACAGT 34188  
 QY 257 GluTyrIleuThrTyrArgValAlaAspGlyValSerValArgGlySerAspGlyThr 276  
 Db 34189 TTAAGTCGTTCTTAACGACCAACCATCTCAAGAAATGACATTAACAAATGCG---GGTAAG 34245  
 QY 277 ThrThrArgThrArgArgProPhePheAsnTyrLeuGluGluGluLeuGluArgGarg 296  
 Db 34246 GTGACACGCACTCAAGAGATATTTTGACTATCTCAACGCCAAATTAATTCCTCGCGCT 34305  
 QY 297 ValProValAlaProGluLeuProPheGluPheAsnLeuGlyTyrValGlyTyrLeuGly 316  
 Db 34306 TGGCCCA---TCTGATGAGCTTAACCTTTTGATTTTAACGTGCTGTTGGGCTACTTTGGT 34362  
 QY 317 TyrGluLeuLeuYsaIagIuThrThrGlyAspProAlaHisArgSerProHisProAspAla 336  
 Db 34363 TATGAAGTGAAGCAAAATCCGGGCTCGGGTTAGAGATCCCTCATTTATTTCCGATGCT 34422  
 QY 337 AlaPheLeuPheAlaAspArgAlaIleAlaLeuAspHisGlnGluGlyCysCysTyrLeu 356  
 Db 34423 ATCTGATTTTAAGCCGATCCCTGATGCTTTTGACCAACCAAGAACCACTTACTTA 34482  
 QY 357 LeuAlaLeuAspArgArgGlyHisAspAspGlyValaArgAlaThrLeuArgGluThrAla 376  
 Db 34483 GTTTACTCTGCGCCCTGTGGGTGAACACGACCGACCCAGCCCTGGTTGAACCTACTGAA 34542  
 QY 377 GluThrIleuThrGlyLeuAlaValaArgAlaProAlaGluProThrProAlaMetValPhe 396  
 Db 34543 AAACGCTCGCGGACACTC-----CCGCTCTCTCTCCCTCGTTCCCT 34584  
 QY 397 GlyTyrIleProGluAlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHisAspLysAsp 416  
 Db 34585 AGTACTACATCTGACCCCGTACTTTC-----CGTTAAGTCGGTACTACAG 34632  
 QY 417 AlaTyrLeuLysArgIleAspGluCysLeuLysGluIleArgAsnGlyGlySerTyrGlu 436  
 Db 34633 ACCTATTAAGATATATTTTACATGCTCAAGAAATTTGTAGGGGGAACCTTACCAA 34692  
 QY 437 IleCysLeuThrAsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTyrSer 456  
 Db 34693 GTTTGCTGACTATGATTAATGACACAGACAAATGCTATGCATTTGAATTTTATCCG 34752  
 QY 457 AlaLeuArgAlaIleSerProValProTyrGlyAlaLeuLeuGluPheProGluLeuSer 476  
 Db 34753 CGATTACGCCAAGTTATCCGGCTCTACTGCTTTTCTACGGTTTGGTGAGATGCGC 34812  
 QY 477 ValIleuSerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGlySer 496  
 Db 34813 ATGCGCTGTTCTTCCCGAGCCATTTCTGCAAAATGACCGTCAAGGCTGGGTAGAAACC 34872  
 QY 497 LysProIleLysGlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuArg 516  
 Db 34873 AAGCCGATCAAGGAAGTGTAGCCCGGAGAGACTCCGAAAGAAGTTTTTACTCCG 34932  
 QY 517 AlaAspLeuAlaGlyArgGlyLysAspArgAlaGluAsnLeuMetIleValAspLeuVal 536  
 Db 34933 GAAAGTTTGGGTGACGCGCAAAAAGCCGATCCGAAATTAATGATTGTTGATTATTA 34992  
 QY 537 ArgAsnAspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGlu 556  
 Db 34993 CGCAAGCATTTAAGCGCAGTTTCCAAAGTGGGAGATCCATGTTCCCAATTAATGAT 35052  
 QY 557 ValGluThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgPro 576  
 Db 35053 GTGAAAGATTATCCAAACGCGTCACTTGTCAAGCATATCCGTGTTATTAATCCCC 35112  
 QY 577 GlyThrSerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAla 596  
 Db 35113 GACATGACGCAACCAAGTGCATTTGCTGGCATTTCCGGCGGCTCGATGACAGGTGCG 35172

QY 597 ProLysLysArgThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTyr 616  
 Db 35173 CTTAACTGAGAAACGATGAGATTTATTGATCACTTGACCAAGAGCGCGGAGTTTAT 35232  
 QY 617 SerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArg 636  
 Db 35233 TCAGGGCGCATCGTCTTTTGGGTGAACCGGTACGCGCACTTGAAATATTGTCATCCGT 35292  
 QY 637 ThrIleValLeuAlaAspGlyGluGluPheGlyValGlyGlyAlaIleValSerLeu 656  
 Db 35293 ACCGCGGTACTGACACCAATCAAACTTCTATGTGATGGGAGGTGCGATTTGCGACAC 35352  
 QY 657 SerAspGluGluGluGluPheThrGluThrValIleValAlaArgAlaMetValThrAla 676  
 Db 35353 TCTGATGTGAAATGGAATTTCAAGAACTATCTCAAGCTTAATTCGCTAATTCAGGCG 35412  
 QY 677 Leu 677  
 Db 35413 ATG 35415  
 RESULT 13  
 ST575648 29132 bp DNA linear BCT 23-JAN-2004  
 LOCUS  
 DEFINITION  
 (aurABCDPFGHI genes), strain HKI-22.  
 ACCESSION  
 AJ575648.1 GI:41322765  
 VERSION  
 AJ575648  
 KEYWORDS  
 acyl-coA ligase; aurA gene; aurB gene; aurC gene; aurD gene; aurE gene; aurF gene; aurG gene; aurH gene; aurI gene; cytochrome P450; N-oxidase; O-methyl transferase; PABA synthase; polyketide synthase type I; transcriptional regulator.  
 SOURCE  
 Streptomyces thioIuteus  
 Streptomyces thioIuteus  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 ORGANISM  
 Streptomyces thioIuteus  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 REFERENCE  
 He,J. and Hertweck,C.  
 TITLE  
 Iteration as programmed event during polyketide assembly; molecular analysis of the aureochin biosynthesis gene cluster  
 JOURNAL  
 Chem. Biol. 10 (12), 1225-1232 (2003)  
 PUBMED  
 14700630  
 REFERENCE  
 2 (bases 1 to 29132)  
 Hertweck,C.  
 TITLE  
 Direct Submision  
 JOURNAL  
 Submitted (07-JUN-2003) Hertweck C., Bioorganic Synthesis, Hans-Knoell-Institute, Beutenbergstr. 11a, Jena, 07745, GERMANY  
 FEATURES  
 source  
 1..29132  
 /organism="Streptomyces thioIuteus"  
 /mol\_type="genomic DNA"  
 /strain="HKI-227"  
 /db\_xref="taxon:66431"  
 2154..2969  
 /gene="aurD"  
 2154..2969  
 /gene="aurD"  
 /function="transcriptional regulator"  
 /codon\_start=1  
 /transl\_table=11  
 /product="AurD protein"  
 /protein\_id="CAE02599.1"  
 /db\_xref="GI:41322766"  
 /translation="WVNAVVKLPSPLSRNILCPVAVSTEGRLPIGPAPORAVSALLLAPGVVGAERLTAIWPAGPPGNAVSLHSYVSHRQLLEDFDPPWGNRLTLREPOGYLAIVAEVRVDARFPEGRLREGRLQLEGRYPESQVIRACLELMRGSPYAELEQDYEPAVCEAAERLERLATLLEALWEALWGHNDVYTAIAELAAVASARPTBERLQWLMLVAICEVROAEAVVAVNRIRKALAOELGVDPGRRLRLPRLRLDSVGRAICQS"  
 complement(3029..4546)  
 /gene="aurE"  
 complement(3029..4546)  
 /gene="aurE"  
 /codon\_start=1



[illegible][illegible]







Qy	641	AlaApGlyGlnAlaGluPheGlyValGlyValAlaIleValSerLeuSerAaspGlnGlu	660	/organism="Streptomyces avermiltillis MA-4680"
Db	13689	ACCGGTACGGGACGACGATCGGACCGGGGGCCCATCTCTCGACTCGACCCGGAC	13748	/mol_type="genomic DNA"
Qy	661	GluGluPheTherGluThrValValAlaArgAlaMetValThrAla-----Leu	677	/strain="MA-4680"
Db	13749	GAGGAGTTTCAGCAGGTGCTCAAGGCGACGCCCTGTGTCGGCGGCGACGCGCTGCTG	13808	/db_xref="taxon:227882"
Qy	678	AapGlySerAlaValAla	683	/note="This strain is also named as strain: ATCC 31267, NCIMB 12804 or NRRL 8165.-synonym: Streptomyces avermectinus"
Db	13809	GAGGGGGCCGCGAGGCC	13826	208..921
RESULT 14				/gene="orf-1"
AP005025/C				208..921
LOCUS	AP005025	298300 bp	DNA	/gene="orf-1"
DEFINITION	Streptomyces avermiltillis genomic DNA, complete genome, section	5/30.		/note="SAV944 not involved in biosynthesis of avermectin"
ACCESSION	AP005025	BA000030		/codon_start=1
VERSION	AP005025.1	GI:29604584		/transl_table=11
KEYWORDS	Streptomyces avermiltillis MA-4680			/product="reductase"
SOURCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			/protein_id="BAC68654.1"
ORGANISM	Streptomycinae; Streptomycetaceae; Streptomyces.			/db_xref="GI:29604585"
REFERENCE	1			/translation="MCVDRDGRGAGVTADLTARGRARAAMPVADVDDGAMRPAER
AUTHORS	Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H.,			GAHIVTASAAAFQPTRAVPAYATSKAAVLMISECLRAEFAEFGVGVSVCPGFRVS
TITLE	Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,			FASAMHFAGVPRLEQERLRALPAGRCSEAKYAAAVLRSVARDSAVVTATYABARLSRL
TITLE	Kikuchi, H., Shiba, T., Sakaki, Y., and Hattori, M.			MSRFTPLRAAAVARMDDPS"
TITLE	Genome sequence of an industrial microorganism Streptomyces			979..2217
TITLE	avermiltillis: deducing the ability of producing secondary			/gene="aveB1"
TITLE	metabolites			979..2217
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	98 (21)	12215-12220 (2001)	/gene="aveB1"
PUBLISHED	21477403			/note="SAV945
REFERENCE	11572948			avermectin gene cluster"
AUTHORS	2			/codon_start=1
TITLE	Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H.,			/transl_table=11
TITLE	Shiba, T., Sakaki, Y., Hattori, M. and Omura, S.			/product="glycosyl transferase"
TITLE	Complete genome sequence and comparative analysis of the industrial			/protein_id="BAC68655.1"
TITLE	microorganism Streptomyces avermiltillis			/db_xref="GI:29604586"
JOURNAL	Nat. Biotechnol.	21 (5)	526-531 (2003)	/translation="MSDHELPMASAPFMGHVPPSLAVALAEIYVRGHHTVETGAEEMADA
PUBLISHED	22608306			VASVGDPLRYESAPEGVDVMTAEENALPMTLYDGMGLSVSEVHGVDPL
REFERENCE	12692562			VAYDATSINVRVLAASGRAPMTVYPLPASNCRPSTQSVLDPSAOGVAPPPRS
AUTHORS	3 (bases 1 to 298300)			EBMELRGALVPRLAELIVSRGITPEVDPLSGDEPNVCLPRAFOYAGTDFDRF
TITLE	Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,			LOKGVDPVLRPLPNNVEYRWVPHAVLEHARALVTGCTGSVMEALHAGCPVLVP
TITLE	Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,			LSRDQVTRRIYAEIGLGMVQPBVEVATTLARHVLDIISDAITRQVRQMRATVEA
TITLE	Kushida, N., Director-General of Biotechnology Center, Shiba, T.,			GALRAADTERRFLRTRRH"
TITLE	Sakaki, Y. and Hattori, M.			complement(2123..3190)
TITLE	Direct Submission			/gene="aveB11"
TITLE	Submitted (29-MAR-2002) Director-General of Biotechnology Center,			complement(2123..3190)
TITLE	National Institute of Technology and Evaluation, Biotechnology			/gene="aveB11"
TITLE	Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan			/note="SAV946
TITLE	(E-mail: bio@nitech.go.jp, URL: http://www.bio.nitech.go.jp/,			avermectin gene cluster"
TITLE	Tel:81-3-3481-1933, Fax:81-3-3481-8424)			/codon_start=1
TITLE	This work was done in collaboration with Haruo Ikeda (*1), Jun			/transl_table=11
TITLE	Ishikawa (*2), Akiharu Hanamoto (*3), Chigusa Takahashi (*3), Jun			/product="dtdp-glucose 4,6-dehydratase"
TITLE	Shinose (*3), Hiroshi Horikawa (*4), Hidekazu Nakazawa (*4), Tomomi			/protein_id="BAC68656.1"
TITLE	Osonoe (*4), Norihiro Kushida (*4), Hisashi Kikuchi (*4), Tadayoshi			/db_xref="GI:29604587"
TITLE	Shiba (*5), Yoshiyuki Sakaki (*6, *7), Masahira Hattori (*1, *7)			/translation="MTGAGGRTGSHFVRRLTGAPAPFGAENVVLDKTYAGRLNLT
TITLE	and Saeoshi Omura (*1, *3).			APVLSPSITFPHGTCIGPLVADIMDSQDMVHFAASSHVDRSVAADAEFRTNVLG
TITLE	Final finishing process and all annotation were done by H. Ikeda			THTLTAAVDAVDRFVY1STDEYVGSIDSGWTDAPLEPSPYSASASVADLLARD
TITLE	and J. Ishikawa.			FRRTGHPVITIRCSNNYGPQFPKLLPRFTVHLNGTKVPLYGSGENVRDMLHVD
TITLE	*1 Kitasato Institute for Life Sciences, Kitasato University			HCRGIALVAERGPGEIYHIGGTELSRRELTPARLLDLGVMSWVEPTVDRKGHNR
TITLE	*2 National Institute of Infectious Diseases			YSLDSKLSAELRGVAPVPEERGLAQVTQWYVENRTWMEPLTARBELPVDSASAGET
TITLE	*3 The Kitasato Institute			AASRLPAGRRPPRPWPASAA"
TITLE	*4 National Institute of Technology and Evaluation			complement(3187..4086)
TITLE	*5 School of Science, Kitasato University			/gene="aveB11"
TITLE	*6 Institute of Medical Science, University of Tokyo			complement(3187..4086)
TITLE	*7 RIKEN, Genomic Sciences Center			/gene="aveB11"
TITLE	Following url is also available.			/note="SAV947
TITLE	http://avermiltillis.jis.kitasato-u.ac.jp.			avermectin gene cluster"
TITLE	Location/Qualifiers			/codon_start=1
TITLE	1..298300			/transl_table=11
FEATURES				/product="glucose-1-phosphate thymidyltransferase"
source				/protein_id="BAC68657.1"
source				/db_xref="GI:29604588"
source				/translation="MKGIVLAGCGRSLRYPLTRALSOLLPVYDKPMYYPLSVLMG
source				GKIDLIVSSPDHLEQFRRLADGSRILGIDYIAQQRPGGIAEAFLLGADFTGDDV







Db 267013 GGCCAGCGCTTCGACTACTGAAACGAGCTGGTGAACCGCTGAGTGAT---GCCAC 266957  
 Qy 301 OGluLeuProPheGluPheAenLeuGlyTyrValGlyTyrLeuGlyTyrGluLeuVal 321  
 Db 266956 GGAATCCCGCTTCACTTACCGGCGGTATGTGGCTACTTCCGCTACGAGGGAAGGC 266897  
 Qy 321 aglUtrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPheAl 341  
 Db 266896 CGACTGGGCTTCCCTCAACCGTATCCGGGCCCAACTCCGACCCCTGCTGCTTCCG 266837  
 Qy 341 aAspArgAlaAlaLeuAenPheHisGlnGlyCysCysTyrLeuLeuAlaLeu---- 359  
 Db 266836 CGACCGGCTGATCCCATGAGCATGAGCAGAGCAGACCTACCGCTGCTGCTCGCGCA 266777  
 Qy 360 ----AspArgArgGlyHisAspArgGlyAlaArgAlaTyrLeuArgGluThrAlaGluThr 378  
 Db 266776 GAACACCCGGGAGGGCCGACAGCAAGCGGCC--GACTGGCTCAACCGCGGATGAGCCA 266720  
 Qy 378 rLeuThrGlyLeuAlaValArgAlaPro---AlaGluProThrProAlaMetValPheG 397  
 Db 266719 GCTTACGTTTGTGTCTCCGCGCAAGCCGACCGCCCGCGCGCCGACG----- 266671  
 Qy 397 ylleProGluAlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHisAspArgAla 417  
 Db 266670 -GCACCCCGCTCGATGCGGTGAGCGGTGCTGCTCGC-----GACCGTACGAC 266621  
 Qy 417 aTyTyrLeuValArgGlyLeuAspGlyCysLeuValGluTyrAlaArgAsnGlyGluSerTyrGlu 437  
 Db 266620 CTACCTTCGCGACATCGGACCTCGCAAGCTCAAGACCGGACCGCATTCAGAT 266561  
 Qy 437 eCysLeuThrAsmMetValThrAlaProThrGluAlaThrAlaLeuProLeuTyrSerAl 457  
 Db 266560 CTGTCTTACCAACGCGCGCCAGCGCTGCCCGCTACAGCCCTTACGATCCGGGT 266501  
 Qy 457 aLeuArgAlaIleSerProValProTyrGlyAlaLeuLeuGluPheProGluLeuSerVa 477  
 Db 266500 ACTGCGTCTGCTCGACCGCGCGCTTACCGCGCTTACCTGCGGTGATCCTCGATGT 266441  
 Qy 477 lleuSerAlaSerProGluValPheLeuThrIleGlyAlaAspGlyGlyValGluSerTyr 497  
 Db 266440 GGCGGCTCTCCCTCCGAGGCTTTTCGCGATCACCAGCGGACGCGCTCCCGGAGCGAC 266381  
 Qy 497 sProIleValGlyThrArgProArgGlyGlyThrAlaGluGluAspGluValArgAla 517  
 Db 266380 ACCGCTCAAGGGACCGCGCCCGCGCGAGCGCGCGGAGAGAGACCCCGCTACCGGA 266321  
 Qy 517 aAspLeuAlaGlyArgGluValAspArgAlaGluAsnLeuMetIleValAspLeuValAr 537  
 Db 266320 CGGCTCACCAAGAGAGACCAAGACCCGCGGAGAACCTGTATGTCGACCTGCTCCG 266261  
 Qy 537 gAsnAspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluVa 557  
 Db 266260 CAACGACCTGGCGCGGCTGCGGAGCCGAGACCGGTGAAAGTCAACCGCTGATGGCGAC 266201  
 Qy 557 lGluThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyValArgLeuArgProGl 577  
 Db 266200 CGAGACGTGCGCACCGTGCACCAAGTGTGCTTCCACCGTGAAGGGCGGCTGCGGAGG 266141  
 Qy 577 yThrSerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaPr 597  
 Db 266140 CATCGGAGCGGTGAGCTGCTTACGTCTTCCCGGAGTTCGCTGACCGGGGCGCC 266081  
 Qy 597 oLysValValArgThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTyrSe 617  
 Db 266080 GAAGACGTGCGCACATGATGATCATCGCTGCGAGAACCGAGCCCGCGGGGTGACTC 266021  
 Qy 617 rGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgTh 637  
 Db 266020 CGAGGACCATGCGCTTCCGCTGCGAGTGGGGCGCGGACCTCCCATCCGACGCGCAC 265961  
 Qy 637 rIleValIleuAlaAspGlyGlnAlaGluPheGlyValGlyValIleValSerLeuSe 657  
 Db 265960 GGCGGTGTTACCGAGCGGAGATGCACTCGGCGCGGGCGGTGCAATCCTCTCGGCTC 265901

Qy 657 rAspGlnGluGluPheThrGluThrValValIleValArgAlaMetValThrAlaLe 677  
 Db 265900 CGATCCCGTGGCGGAGTACACAGATGCTGTAAAGCGCGCACCGATGGCGCC-- 265843  
 Qy 677 uAspGlySerAlaValAlaGlyAlaArg 686  
 Db 265842 -CACCGGACCGGATCCCGGCTCGCG 265816  
 RESULT 15  
 AP005217  
 LOCUS  
 DEFINITION  
 411  
 AP005217 300750 bp DNA linear BCT 24-JUN-2003  
 Corynebacterium efficiens YS-314 DNA, complete genome, section  
 411  
 AP005217 BA000035  
 AP005217.1 GI:23492722  
 SOURCE  
 ORGANISM  
 Corynebacterium efficiens YS-314  
 Corynebacterium efficiens YS-314  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
 1  
 Nishio, Y., Nakamura, Y., Kawarabayashi, Y., Usuda, Y., Kimura, E.,  
 Sugimoto, S., Matsui, K., Yamagishi, A., Kikuchi, H., Ikeo, K. and  
 Gojobori, T.  
 Comparative complete genome sequence analysis of the amino acid  
 replacements responsible for the thermostability of Corynebacterium  
 efficiens  
 Genome Res. 13 (7), 1572-1579 (2003)  
 22723752  
 12840036  
 2 (bases 1 to 300750)  
 Kawarabayashi, Y., Yamazaki, J., Hino, Y., Kikuchi, H. and  
 Director-General of Biotechnology Center.  
 Direct Submission  
 Submitted (17-MAY-2002) Director-General of Biotechnology Center,  
 National Institute of Technology and Evaluation, Biotechnology Center,  
 Center; Nishihara 2-49-10, Shibuya-ku, Tokyo 151-0066, Japan  
 (E-mail: bio@nitech.go.jp, Tel:81-3-3481-1933, Fax:81-3-3481-8424)  
 Kawarabayashi, Y. is officially affiliated with the National  
 Institute of Advanced Industrial Science and Technology, Tsukuba,  
 Ibaraki, 305-8566 Japan  
 Nakamura, Y., Ikeo, K., Suzuki, M. and Mashima, J. are at the  
 National Institute of Genetics, Mishima, Shizuoka, 411-8540 Japan  
 Itoh, T. is at the Japan Biological Information Research Center,  
 Koto-ku, Tokyo, 135-0064 Japan  
 Yamagishi, A. is at Tokyo University of Pharmacy and Life Science,  
 Hachioji, Tokyo, 192-0392 Japan  
 Nishio, Y., Usuda, Y. and Sugimoto, S. are at the Ajinomoto Co.,  
 Inc., Kawasaki, Kanagawa, 210-8681 Japan  
 The other authors are at the National Institute of Technology and  
 Evaluation, Shibuya-ku, Tokyo, 151-0066 Japan.  
 Location/Qualifiers  
 1..300750  
 /organism="Corynebacterium efficiens YS-314"  
 /mol\_type="genomic DNA"  
 /strain="YS-314"  
 /db\_xref="taxon:196164"  
 252..620  
 /note="CE0885"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC17695.1"  
 /db\_xref="GI:23492723"  
 /translation="MDQAPSGGLDNCVGYIPYDPVPVPLLPVYGTSGNRKIVLDFN  
 HFWLNKPRILITPLMSGGCLFGWVIGLFTWHVHVVCVFCBFPDGSRTLAAC  
 LTHASTKPKCQGGIVANG"  
 529..1936  
 /gene="rRNA\_16S"  
 529..1936  
 /gene="rRNA\_16S"  
 /product="16S ribosomal RNA"



```
CDS
/note="CE:RNA01"
/complement (2042..2215)
/note="CE0886"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC17696.1"
/db_xref="GI:23492724"
/translacion="MSSQTPNSMPTYYHHHLDGVADPWMMIHPIITTCWMOQHPGSGNT
NMPHESPMWVY"
2399..5481
/gene="trna_23s"
2399..5481
/gene="trna_23s"
2399..5481
/gene="trna_23s"
/product="23s ribosomal RNA"
/note="CE:RNA02"
5606..5725
/gene="trna_5S"
5606..5725
/gene="trna_5S"
5606..5725
/product="5S ribosomal RNA"
/note="CE:RNA03"
5925..7088
/note="CE0889, similar to AX063829-1|CAC25156.1| percent
identity: 79 in 387 aa"
/codon_start=1
/transl_table=11
/product="putative aminotransferase"
/protein_id="BAC17699.1"
/db_xref="GI:23492727"
/translacion="MRNKYVERLRPFGETTIFATWTORANEAGAINLGGCFDEDEGPA
PMLEISOQILGNNQYSGGSDLGRTAVASRHRFKLDYDPDTEVLIYVATGTAI
SASVLGLVERGDEVIILEPYDAYAAAIAGATRYAVPLKEEDNTWVLDITGATNAV
TKTSMIIVNSPHNPTGVSFKSLSLAGIARAYDLVLSDEVEHLLPDKXSHYAV
ASIPGMMDRTITYSSAKTPNTYGMKTGMAIAPILNAVIRAKQMSYVGATPPRPA
VAAVAHKKPWAQKMRKGLONKSDIITDALTOGLVHSHSGTYFVAADIGEDGAEF
CFPLIDRVGVAAIPVOAFVDPKPEQSSKVFACQKQDITLREAAQRLRAAGSL"
complement (7180..7821)
/note="CE0890"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC17700.1"
/db_xref="GI:23492728"
/translacion="MTHSFSDVDEAVARKNNEFLRDTKRLQISAPFFGGLVLAIGVL
YLDAGVAGMWMLIVMGIMALSFMIPIPLPGVGSASISLVNTYVPLVAVVAENPR
DVYLLAVNTNVDETIIPRWGLAVRTINRLGVDRKLGERSVSVANSGRRTIRDOHW
DEISPMPLTWGITDADVVRRAEKTIIPHDLMNKLKRNRLDEVKTRMLDLV"
complement (7822..9495)
/note="CE0891, similar to AL583924-93|CAC31112.1| percent
identity: 75 in 547 aa"
/codon_start=1
/transl_table=11
/product="putative DNA helicase"
/protein_id="BAC17701.1"
/db_xref="GI:23492729"
/translacion="MAFGDGLIVOSDKTVLLETGDLAGEARALAPAELEBERAHEH
VHTYRITPLALNNARAAGHAEQVVDVLEBYSFPPOPLVNAETMGRVGVNRHK
HPNAGLILBESSEPAIIVETSRHKIKPMIGQOUDPETIAVHPSERGLKQELKTMF
AEDLAGVDEAHPIALSTVEDWSLRDYOQVADSFWEQSSVVVLPFGAGTWMGA
ASMARAOATLLIVNTVAGRWKMDLRRTLRTTBEGISGERKEIPRVIATYOV
VTRRTKGEYKALEIPSRDWGLIYDEVHLLPAPVPMSTSDLOSRRRLGILATVED
GREGVDSLIGPRYDAPWKDLEAGFIATADCEVETMTDAERNVYVATFADNRV
LATATHTKTRVVRVRIIEOHOGQPTLIIGATLDLELGAEPDAIYDQGTSNKREEL
PFRFRSGELNVVSKYANFSIDLPRAAVAIVSGTFGFSROEBAQRLGRLPKADNG
EAHFYVSVSRDITDTEYAAHRQRFLEAQGAARIIDADDLFPITEKET"
complement (9548..11851)
/note="CE0892, similar to AE006976-11|AAK45126.1| percent
identity: 36 in 741 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC17702.1"
```

```
/db_xref="GI:23492730"
/translacion="MPCPEFSQIRQTKVTSAPKLGDIIMKDSVPVPTLTGWLQOLD
OLSGILRNRPDPLPLPGLGSLARLORABIOAVHKLSELEAVLEAANAGAEI
NPITAPEYVERLHRLRHSAAVPTAEQIGELANRLROLALVGRGRFTILPEMTALP
TNWOLLPEHGRSLTRGESQSIEALPARHKITVTLVTSNGLGTRDAISADPARP
IPQLIKAGLSRVDEQIVKLPGVRRVLEGEQDITAMRVVPEPSAVPGADLGIAGL
EVARLQRMILDALSTAPASTALEKAGVAMVTRLTREIQVDQVAVVSLCVSSGL
RGVDPDPLPADDDGSDYIAPTPVADBEWLOATLAEQLAHLMRGMTQTYAAMVGEAD
KKKPIHLSPASRDALPETRAMILOSILTRVEEDLEADLFPHHPLASRIIPRETRH
LLEBARWGAVAGGVTSAPVRLTOAPRGVIEIADIVAVRAPKVDVYFIVQADVTY
MYPGQLPEMKQITIQIADLBSPGLASVYRISBASLRHAMDGLTATEIEDITQSHA
TGLPOSVALILODIARRGTLRGSPALCYLRSDDPALLHSAVEAGALDRLDIAPIVA
VSNASLIEVITLRRAGQVPAEDALGASLDIAPPARVPAEQPDSAGPDESRIRA
AVAAIRREBEVARTGVSDQPTLAVIQAAVRCQRTVTLGFEVDKQGVAVHRTVPLTVNA
GOVDALDESTGAVHRFMLRITEIVD"
11836..12036
/note="CE0893, similar to AL583924-91|CAC31110.1| percent
identity: 50 in 62 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC17703.1"
/db_xref="GI:23492731"
/translacion="MGNMAVEXKGYVDPAMPDHPDSEGHVTELIAPYAGASPEWD
DMEFVPARTIGVHPYTRINR"
complement (12229..12822)
/note="CE0894, similar to AX133781-1|CAC38957.1| percent
identity: 71 in 195 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC17704.1"
/db_xref="GI:23492732"
/translacion="MLFMGRHSAKRATPEVKMAATTVAIRGAANAAPAPASAAPSDW
DLANCESGNAINTGNGYHGLDPSSTWMAVYGDDEAPAPAYATREQQAVALAERT
LAGQMGAMPACSSKGLNSAPTRPNLNPAPAPAPAPAPARATVDTNTNPVVGSS
DNTQVGIYDAVSGTLAQGVQVPAEIQAHVYAFGR"
13342..13725
/note="CE0895, similar to AE006977-4|AAK45135.1| percent
identity: 52 in 127 aa"
/codon_start=1
/transl_table=11
/product="putative cold shock protein"
/protein_id="BAC17705.1"
/db_xref="GI:23492733"
/translacion="MPGVKQVYDPERGFGVSNBGEPCFYGXOVLPKGYTELHOG
KIEFPFAPGRKGPQLARVAVLETPRRRPQHTYKPEBLNGIISDMVTMLEQVQALLK
GHYPRKVGEOYAKILIRAVAKELA"
complement (13881..14408)
/note="CE0896"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC17706.1"
/db_xref="GI:23492734"
/translacion="MASRKYKKNFIQIATLLVAVVVVISVVFQNMNNRNEPEPE
TVSIYASSPAGIEVPEYPSICEGVCACNEVPTLEVGDELRLTIPEIIRHDMWYL
LIYVDPAANDREFYHTSYETTEVYTPGSVDPTVEGGERRLVVEVTSVMIGTDNGE
ETPYTTTWSLATEVO"
complement (14424..15254)
/note="CE0897, similar to AE001874-1|AAF09704.1| percent
identity: 40 in 267 aa"
/codon_start=1
/transl_table=11
/product="putative glutamine cyclotransferase"
```

## Alignment Scores:

Pred. No.:	1.6e-64	Length:	300750
Score:	1541.00	Matches:	342
Percent Similarity:	60.71%	Conservative:	69
Best Local Similarity:	50.52%	Mismatches:	206
Query Match:	43.13%	Indels:	60
DB:	1	Gaps:	13



US-10-089-514-2 (1-686) x AP005217 (1-300750)

QY 1 MetArgThrIleuLeuIleAspSerPheThrHisAsnLeuPheGluTyrIle 20  
184888 GTGGCGCTACTGATCATCGCAATTAATGATTTCTTCCCTTCAACCTCGCACGATATGTC 184947  
DB 184948 GAGAGGCTCACCGGTGGGACCGACGCGGTGGTCCCAATGATGCG-----CAATC 184998  
QY 21 GlyIuAlaThrGlyGlnProProValValProAsnAspIleAspTyrPheArgLeu 40  
184948 GAGAGGCTCACCGGTGGGACCGACGCGGTGGTCCCAATGATGCG-----CAATC 184998  
DB 184948 GAGAGGCTCACCGGTGGGACCGACGCGGTGGTCCCAATGATGCG-----CAATC 184998  
QY 41 ProValGluAspPheAspAlaIleValSerProGlySerProAspArgGlu 60  
184999 GACGAGACGCTGTTGACGCGCTCATCATCTCACCGCGGCCCGCCACCGCGGGGTGGC 185058  
DB 184999 GACGAGACGCTGTTGACGCGCTCATCATCTCACCGCGGCCCGCCACCGCGGGGTGGC 185058  
QY 61 ArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValIleGlyVal 80  
185059 GCGGATTTCCGCGACTCCCGGGGTGATCGAACTGCTGCTGCTCCGCTCGGGGGTGC 185118  
DB 185059 GCGGATTTCCGCGACTCCCGGGGTGATCGAACTGCTGCTGCTCCGCTCGGGGGTGC 185118  
QY 81 CysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaProGlu 100  
185119 TGCTGGGACATACGGGACATCGGCTCGCCACGCGCGCGGTGGGCGCGCGCTC 185178  
DB 185119 TGCTGGGACATACGGGACATCGGCTCGCCACGCGCGCGGTGGGCGCGCGCTC 185178  
QY 101 ProMetHisGlyArgValSerGluValArgHisThrGlyIuAspValPheArgGlyLeu 120  
185179 CCGGTGACCGGTGACGTCACCGCATCCACACGACGCGGTCCGAGCTTTTCGACGCCATC 185238  
DB 185179 CCGGTGACCGGTGACGTCACCGCATCCACACGACGCGGTCCGAGCTTTTCGACGCCATC 185238  
QY 121 ProSerProPheThrAlaValArgTyrHisSerLeuAlaIleThrAspLeuProAspGlu 140  
185239 CCCCCCAGTTTCACGCGCGCTTACCTCCCTGCTGCGCACCGATCTGCGCGCGGA 185298  
DB 185239 CCCCCCAGTTTCACGCGCGCTTACCTCCCTGCTGCGCACCGATCTGCGCGCGGA 185298  
QY 141 LeuGluProLeuAlaTyrPheSerAspAspGlyValValMetGlyLeuAlaGlnGlyLeu 160  
185299 CTGAGAGTCAACCGGACGACCGGGAGCGGTATGCTATGCGCTGCCACCGGAGCTG 185358  
DB 185299 CTGAGAGTCAACCGGACGACCGGGAGCGGTATGCTATGCGCTGCCACCGGAGCTG 185358  
QY 161 ProLeuTyrGlyValGlnPheHisProGluSerIleGlySerAspPheGlyArgGluIle 180  
185359 CCCCAGTGGGCTGTCAGTTCCACCGGAACTCATCGCGCGGACGTTCCGCGCACCGGATC 185418  
DB 185359 CCCCAGTGGGCTGTCAGTTCCACCGGAACTCATCGCGCGGACGTTCCGCGCACCGGATC 185418  
QY 181 MetAlaAsnPheArgPheLeuAlaLeuAlaHisHisArgAlaArgArgHisGlyValAsp 200  
185419 ATGCGCAATTTTC-----CTGACGCTGGCGCGTCGACGACGCGACCGG----- 185460  
DB 185419 ATGCGCAATTTTC-----CTGACGCTGGCGCGTCGACGACGCGACCGG----- 185460  
QY 201 SerProTyrGluLeuHisValArgArgValAspValLeuProAspAlaGluValArg 220  
185461 -----TGGAAGATCACCGAGCATGTGTGAGACAAGCGTCAACCGCGCGCGGTGTC 185514  
DB 185461 -----TGGAAGATCACCGAGCATGTGTGAGACAAGCGTCAACCGCGCGGTGTC 185514  
QY 221 ArgGlyCysLeuProGlyGluGlyThrThrPheTyrLeuAspSerSerValLeuGlu 240  
185515 GAGACGCTTTCGCCCGGGTGGAGACGCGCTTCTGGCTCGATGATCCG-----CAG 185565  
DB 185515 GAGACGCTTTCGCCCGGGTGGAGACGCGCTTCTGGCTCGATGATCCG-----CAG 185565  
QY 241 GlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGlyTyrLeuThr 260  
185566 GGCACG-----ACCTATATGGGTGATGCCCGCGCGCATGACGAGATCCGAC 185616  
DB 185566 GGCACG-----ACCTATATGGGTGATGCCCGCGCGCATGACGAGATCCGAC 185616  
QY 261 TyrArgValAlaAspGlyValValSerValArgGlySerAspGlyThrThrArgThr 280  
185617 CACCGGGTGGGAGGG----- 185634  
DB 185617 CACCGGGTGGGAGGG----- 185634  
QY 281 ArgArgProPhePheAsnTyrLeuGluGluGlnLeuGluArgArgValProValAla 300  
185635 -----GAGCTGTTTCACTGAGCTGCGTATGATCTGCGTCAACCGC-----GTGGCC 185682  
DB 185635 -----GAGCTGTTTCACTGAGCTGCGTATGATCTGCGTCAACCGC-----GTGGCC 185682  
QY 301 ProGluLeuProPheGluPheAsnLeuGlyTyrValGlyTyrLeuGlyTyrGluLeuLys 320  
185683 CCGGGGGTGTG-----GGTTTCCCGCTGGGGTGGGTGGATACCTCGGATATGAGATGAAG 185736  
DB 185683 CCGGGGGTGTG-----GGTTTCCCGCTGGGGTGGGTGGATACCTCGGATATGAGATGAAG 185736  
QY 321 AlaGluThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe 340  
185737 GCGGAATGCGGGGTGCAATCGCACGCTCATCGATCCCATGCCACCTGATCTTC 185796  
DB 185737 GCGGAATGCGGGGTGCAATCGCACGCTCATCGATCCCATGCCACCTGATCTTC 185796  
QY 341 AlaAspArgAlaIleAlaLeuAspHisGlnGluGlyCysArgTyrLeuLeuAlaLeuAsp 360  
185797 GCGGACCGGGCCATCGCATC-----GAACTCCCGCGGGGTGNGGCTCATGCGGCTCGCT 185850  
DB 185797 GCGGACCGGGCCATCGCATC-----GAACTCCCGCGGGGTGNGGCTCATGCGGCTCGCT 185850  
QY 361 ArgArgGlyHisAspAspGlyAlaArgAlaTyrPheArgGluThrAlaGluThrLeuThr 380  
185851 GACACGAGGAG-----TGTTTCGCGGAGATGACCGCGCGCTGGGG 185892  
DB 185851 GACACGAGGAG-----TGTTTCGCGGAGATGACCGCGCGCTGGGG 185892  
QY 381 GlyLeuAlaValArgAlaProAlaGluProThrProAlaMetValPheGlyIleProGlu 400  
185893 CAGCTG-----CGCCA-----CCCGCT 185910  
DB 185893 CAGCTG-----CGCCA-----CCCGCT 185910  
QY 401 AlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHisAspIleAspValaTyrLeuLys 420  
185911 GCGGCTGCG 185970  
DB 185911 GCGGCTGCG 185970  
QY 421 ArgIleAspGlyCysLeuLysGluIleArgAsnGlyGluSerTyrGluIleCysLeuThr 440  
185971 ATGATCG 186030  
DB 185971 ATGATCG 186030  
QY 441 AsnMetValThrAlaProThrGlyAlaThrAlaLeuProLeuTyrSerAlaArgAla 460  
186031 ACCGACTGCGCGGAGGTGAGGTGATCCCTCGCGCGCTATGCTGCGCGCGCG 186090  
DB 186031 ACCGACTGCGCGGAGGTGAGGTGATCCCTCGCGCGCTATGCTGCGCGCGCGCG 186090  
QY 461 IleSerProValProTyrGlyAlaLeuLeuGluPheProGluLeuSerValLeuSerAla 480  
186091 GCCAACCCACCTCCTATGATGATCTTCGACGCTGGGGAGATGGCGGTGCTGACCTCG 186150  
DB 186091 GCCAACCCACCTCCTATGATGATCTTCGACGCTGGGGAGATGGCGGTGCTGACCTCG 186150  
QY 481 SerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGluSerLysProIleLys 500  
186151 TCGCCGGAAGGTTTCATCACATCGACGCTCGGACGCTGGGATTCGCAACCCATCAAG 186210  
DB 186151 TCGCCGGAAGGTTTCATCACATCGACGCTCGGACGCTGGGATTCGCAACCCATCAAG 186210  
QY 501 GlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuArgAlaAspLeuAla 520  
186211 GGCACCG 186270  
DB 186211 GGCACCG 186270  
QY 521 GlyArgGlyCysAspArgAlaGluAsnMetIleValAspLeuValArgAsnAspLeu 540  
186271 GACAAACCCAAAGACCG 186330  
DB 186271 GACAAACCCAAAGACCG 186330  
QY 541 AsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGluThrTyr 560  
186331 GCCCGGCGGCGCGCAACCGGACGCGTGCAGTGAAGAACCTTTTCAGCGTCGAGCTAC 186390  
DB 186331 GCCCGGCGGCGCGCAACCGGACGCGTGCAGTGAAGAACCTTTTCAGCGTCGAGCTAC 186390  
QY 561 AlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyThrSerThr 580  
186391 GCCACCGTCCACGCTCGTCAGCACCATTAACCGCCACCTGAG-----GTAAGACACCC 186447  
DB 186391 GCCACCGTCCACGCTCGTCAGCACCATTAACCGCCACCTGAG-----GTAAGACACCC 186447  
QY 581 AlaAlaCysValArgAlaIlePheProGlyGlySerMetThrGlyAlaProLysLysArg 600  
186448 ATTGACTGCGTGGCG 186507  
DB 186448 ATTGACTGCGTGGCG 186507  
QY 601 ThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTyrSerGlyAlaLeu 620  
186508 ACCATGAGATCATGACAGACGCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 186567  
DB 186508 ACCATGAGATCATGACAGACGCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 186567  
QY 621 GlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleValLeu 640  
186568 GGTATTTCTCCTTCATGATGGCGCGGTGATGCTGTCATGATGATCGCACCGGCTCTAC 186627  
DB 186568 GGTATTTCTCCTTCATGATGGCGCGGTGATGCTGTCATGATGATCGCACCGGCTCTAC 186627  
QY 641 AlaAspGlyGlnAlaGluPheGlyValGlyGlyAlaIleValSerLeuSerAspGlnGlu 660  
186628 ACCCGCGCGCTGAGTACGAGGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 186687  
DB 186628 ACCCGCGCGCTGAGTACGAGGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 186687  
QY 661 GlnGluPheThrGluThrValValLysValaArgAlaMetValThrAlaLeu 677  
186688 GCAGATGGAGAGATCGCGGTAATGAGAGCCCTGCTGCTGCTC 186738  
DB 186688 GCAGATGGAGAGATCGCGGTAATGAGAGCCCTGCTGCTGCTC 186738

Search completed: October 5, 2005, 15:27:53  
Job time : 10957.6 secs



**This Page Blank (uspio)**











```
QY 521 G1YArgGluYAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAspLeu 540
DB 1561 GGGCGGAGAGAGACCGGGCGAGAACTGATGTCACCTGGTCCGCAACGACTC 1620
QY 541 AsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGluThrTyr 560
DB 1621 AACAGCGTTCGCGCATCGGCTCGTCCAGTCGAGCCCGGCTCTTCGAGGTGAGACTAC 1680
QY 561 AlaProValHisGluLeuValSerThrIleArgGlyArgLeuArgProGlyThrSerThr 580
DB 1681 GCGCCGCTGCACCAAGCTGTGTGACCATCCGGGAGCGGTGGCGCCCGCACGACGACC 1740
QY 581 AlaAlaCysValArgAlaIlePheProGlyGlySerMetThrGlyAlaProLysArg 600
DB 1741 GCGCGCTGCTAGCGCGCGCTTCCCGGCGCTCCATGACCGGCGCGCCCAAGAACGC 1800
QY 601 ThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTyrSerGlyAlaLeu 620
DB 1801 ACCATGGAGATCATCGACCGCTGAGAGAGAGCCCGGGCGGTCTACTCGGGGCGCTC 1860
QY 621 GlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleValLeu 640
DB 1861 GAGTGTCTTCGCTTCAGCGCGCGCGCGCTGACATCGTCAATCCGACCATCGTGTG 1920
QY 641 AlaAspGlyGlnAlaGluPheGlyValGlyValAlaIleValSerLeuSerAspGlnGlu 660
DB 1921 GCGGACGCGGCGGAGGAGTTCGCGCTCGCGGCGGCGATCTGTCTCTCCGACCGAGAG 1980
QY 661 GluGluPheThrGluThrValIleValAlaArgAlaMetValThrAlaLeuAspGlySer 680
DB 1981 GAGGAGTTCACCGAGACCGGTGTAAGGCCGCGGCGCATGTGTACCGGCTCGACGCGAGC 2040
QY 681 AlaValAlaGlyAlaArg 686
DB 2041 GCGGTGCGCGGCGCGCGCA 2058

RESULT 2
AAL50180
ID AAL50180 standard; DNA; 2061 BP.
XX
AC AAL50180;
XX
DT 28-JAN-2003 (first entry)
XX
DE S venezuelae PF1022 substance gene #1.
XX
KW Substance PF1022; biosynthesis; transgenic; choriismic acid; anthelmintic;
KW phenylalanine; p-aminophenylpyruvic acid; pharmaceutical;
KW veterinary drug; gene; ds.
XX
OS Streptomyces venezuelae.
XX
FH Key Location/Qualifiers
FT 1. 2061
FT CDS /*tag= a
FT /product= "PF1022 substance"
XX
XX MO20027244-A1.
XX
XX 03-OCT-2002:
XX
XX PD
XX PF 22-MAR-2002; 2002MO-JP002782.
XX
XX PR 22-MAR-2001; 2001JP-00082227.
XX
XX PA (MEIJ ) MEIJI SEIKA KAISHA LTD.
XX
XX Yanaï K, Sumida N, Watanabe M, Moriya T, Murakami T;
XX
XX WPI; 2003-018934/01.
XX
XX P-PSDB; AAO19563.
XX
XX Novel biosynthesis gene-transferred transformants for producing PF1022
```

```
PT substance derivatives by fermentation, as pharmaceuticals or veterinary
PT drugs with anthelmintic activity.
XX
XX Claim 13; Page 74-79; 116pp; Japanese.
XX
XX The present invention relates to transformants capable of producing
CC PF1022 substance derivatives. These were obtained by transferring a gene
CC participating in the biosynthesis pathway from choriismic acid to p-
CC aminophenylpyruvic acid (biosynthesis gene) into a phenylalanine-
CC requiring host derived from an organism producing the PF1022 substance.
CC The transformants are producing PF1022 substance derivatives by
CC fermentation, for use as pharmaceuticals or veterinary drugs. The present
CC sequence is a substance PF1022 coding sequence from Streptomyces
XX venezuelae
XX
SQ Sequence 2061 BP; 297 A; 842 C; 654 G; 268 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4,65e-245 Length: 2061
Score: 3573.00 Matches: 686
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0

US-10-089-514-2 (1-686) x AAL50180 (1-2061)

QY 1 MetArgThrLeuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGlnTyrIle 20
DB 1 ATGCGCAGCGCTTCTGATCGACCACTGACACTGTTCACTCCACCACTGTTCCAGTACATC 60
QY 21 GlyGluAlaThrGlyGlnProProValValValProAsnAspAlaAspTrpSerArgLeu 40
DB 61 GGGAGAGCCACCGGGGCAACCCCGGTCGTGCGCCCAAGAGCGGCACTGTGCGCGCTG 120
QY 41 ProValGluAspPheAspAlaIleValIleAspProGlyProGlySerProAspArgGlu 60
DB 121 CCGGTGAGAGACTTCGACCGCATCGTGTGTCTCCCGGCGCGGAGCCCGGACCGAGAA 180
QY 61 ArgAspPheGlyIleSerArgAlaIleIleThrAspSerGlyLeuProValLeuGlyVal 80
DB 181 CGGGACTTCGGAATGACCGCGCGGCGGATCACGACGAGCGGCTGCTCCGCGCTC 240
QY 81 CysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaProGlu 100
DB 241 TGCCTCGGCGACAGGCGATCGCCAGCTCTTCGGGGAACCGTCCGCGCGGAGAA 300
QY 101 PromethiGlyArgValSerGluValArgHisThrGlyGluAspValPheArgGlyLeu 120
DB 301 CCCATGACGCGCGGAGTCTCCAGAGTTCGAGCAACCGGCGAGAGAGCTTTCGCGGCTC 360
QY 121 ProSerProPheThrAlaValArgTyrHisSerLeuAlaIleThrAspLeuProAspGlu 140
DB 361 CCTCGCGCTTACCGCGCGTGTGCTACCACTCCCTGCGCGCACGACTCCCGGACGAG 420
QY 141 LeuGluProLeuAlaTrpSerAspAspGlyValIleMetGlyLeuArgHisArgGlyLys 160
DB 421 CTCGAACCCCTCGCGCTGAGCGACGAGGAGTGTCTATGAGGCTGCGGCGGACGCGGAGAG 480
QY 161 ProLeuTrpGlyValAlaGlnPheHisProGlySerIleGlySerAspPheGlyArgGluIle 180
DB 481 CCGCTGTGGGGCGGTCCAGTTCACCGGAGTCACTGAGGAGGAGACTTCGCGCGGAGATC 540
QY 181 MetAlaAsnPheArgAspLeuAlaLeuAlaHisHisArgAlaArgArgHisGlyAlaAsp 200
DB 541 ATGGCCAACTTCGCGGACCTCGCTCGGCCACCAACCGGCGACGCGGCGGCGGCGGAG 600
QY 201 SerProTyrGluLeuHisValArgArgValAspValLeuProAspAlaGluGluValArg 220
DB 601 TCCCGGTACGAATCCACCTGCGCGCGGTGCGAGCTGTCTCCGAGCGCCCAAGATACCC 660
QY 221 ArgGlyCysLeuProGlyGluGlyThrThrPheTrpLeuAspSerSerValLeuGlu 240
```



```

Db      661 CGCGCTGCTGCCCGGCGAGGCGACCACTTGTGCTGAGACAGCTCCGTCTCGAA 720
Qy      241 GlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTyrLeuThr 260
Db      721 GGGGCTCGGGCTTCTTCTTCTTCTCGGGAGACAGCCGGGCGCTCCGCGAGTACTCTACC 780
Qy      261 TyrArgValAlaAspGlyValAlaSerValArgGlySerAspGlyThrThrArgThr 280
Db      781 TACCGGTGCGCGAGCGCGTGTCTCCGTCCGGGCTCGAGCGGACCAAGCCCGAGG 840
Qy      281 ArgArgProPhePheSerTyrLeuGluGluGluLeuGluArgArgValProValAla 300
Db      841 CGGCGCCCTTCTTCAACTACCTGAGAGACAGCTGAAAGCGGAGCGGGTCCCGTCCGC 900
Qy      301 ProGluLeuProPheGluPheAsnLeuGlyTyrValGlyTyrLeuGlyTyrGluLeuLys 320
Db      901 CCGGAAGTGGCTTTCAGATTCAACTCGGCTACGCTAGGCTACTCTCGGCTACGAGCTGAG 960
Qy      321 AlaGluThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe 340
Db      961 GCGGAGACCAACCGGCAACCCGCGCACCGGTCGCCGACCCGACCGCGGCTTCTTTC 1020
Qy      341 AlaAspArgAlaAlaAlaLeuAspHisGluGluGlyCysGlyTyrLeuLeuAlaLeuAsp 360
Db      1021 GCGGACCGGCGCATCGCTCGACCAACAGAAAGGCTGTGCTTACCTGTGCTGCGCTCGAC 1080
Qy      361 ArgArgGlyHisAspAspGlyAlaArgAlaTyrLeuArgGlyThrAlaGluThrLeuThr 380
Db      1081 CGCGGGGGCCACGACGAGCGCGCGCTGCTGCGGAGCGGCGGAGCCCTCACC 1140
Qy      381 GlyLeuAlaValAlaArgAlaProAlaGluProThrProAlaMetValPheGlyLeuProGlu 400
Db      1141 GGGCTGCGCGTCCGCGCCCGCGCGGAGCGGACCCCGGCATGCTTCCGGAGTCCCGGAG 1200
Qy      401 AlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHisAspLysAspAlaTyrLeuLys 420
Db      1201 GCGCGCGCGGCTTCCGCGCCCTGCGCGCGCGCACGACGACGACGCTTACCTCAAG 1260
Qy      421 Arg1LeAspGluCysLeuLysGluLeuArgAsnGlyGlySerTyrGluLeuCysLeuThr 440
Db      1261 CGCATCGACGAGTCCCTCAAGAGATCCGCAAGCGGAGTGTGTATCGATCTGCTGACC 1320
Qy      441 AsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuArgAla 460
Db      1321 AACATGTCACCGCGCGGACCGAGCGGAGCGGCTCCGCTTACTCCGGCTGCGCGCC 1380
Qy      461 IleSerProValProTyrGlyAlaLeuLeuGluPheProGluLeuSerValLeuSerAla 480
Db      1381 ATCAGCCCGTCCGTAAGCGCGCTGTGAGTCTCCGAACTGTCTGCTGAGCGCC 1440
Qy      481 SerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGluSerLysProIleLys 500
Db      1441 TCGCCCGAGGCGTTCCTCAAGATCGGCGCGGAGCGGCGGCTGAGTCCAAAGCCCATCAAG 1500
Qy      501 GlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuArgAlaAspLeuAla 520
Db      1501 GGGACCCGCGCCCGGCGGCGACCGCGGAGAGAGAGAGGCGCTCCGCGCGCATCGGCC 1560
Qy      521 GlyArgGluLysAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAlaPhe 540
Db      1561 GGCCTGGAGAGAGACCGGCGCGAGAACCTGTATATGTGTCACTGTGTCCGCAACGACCTC 1620
Qy      541 AsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGluThrTyr 560
Db      1621 AACAGGCTCGCGCGATCGGCTCCGTCCACGTGCCCCGGCTTCCAGGTGAGACTTAC 1680
Qy      561 AlaProValHisGluLeuValSerThrIleArgGlyArgLeuArgProGlyThrSerThr 580
Db      1681 GGGCCCGTGCACAGCTGTGTGACATCCGGGAGAGCGGTGGCGCCGCGGACCAACACAC 1740
Qy      581 AlaAlaCysValAlaArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLysLysArg 600
Db      1741 GCGGCTGTGTACGCGCGCTTCCCGGGGCTTCATATGACCGCGCGCCCAAGAGCGC 1800

```

```

Qy      601 ThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTyrSerGlyAlaLeu 620
Db      1801 ACCATGAGATCATCGACCGCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Qy      621 GlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleValLeu 640
Db      1861 GGATGGTTCGCTTCAGCGCGCGCGCGGACCTCAGATCGTATCGGACCATCGTGTG 1920
Qy      641 AlaAspGlyGluAlaGluPheGlyValGlyGlyAlaIleValSerLeuSerAspGluGlu 660
Db      1921 GCGGAGCGGCGAGCGGAGTTCGCGCGCGGCGGAGATGTGTCTTCCAGACGAGAG 1980
Qy      661 GluGluPheThrGluThrValIleLysAlaArgAlaMetValThrAlaLeuAspGlySer 680
Db      1981 GAGAGATTACCGAGACCGTGTAAAGCCCGCGCATGTATACCGCTTCAACGCGACG 2040
Qy      681 AlaValAlaGlyAlaArg 686
Db      2041 GCGGTGCGGCGCGCGA 2058

RESULT 3
ABZ69799
ID ABZ69799 standard; DNA; 3305 BP.
XX
AC ABZ69799;
XX
DT 08-APR-2003 (first entry)
XX
DE plasmid papABC.
XX
KW DHFR; translation; orthogonal RNA; O-cRNA; O-RS; CAT; TYRS;
KW orthogonal aminoacyl RNA synthetase; unnatural amino acid;
KW chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;
KW biosynthesis; p-aminophenylalanine; paf; chorismate.
OS Synthetic.
XX
PN WO200285923-A2.
XX
PD 31-OCT-2002.
XX
PF 19-APR-2002; 2002WO-US012465.
XX
PR 19-APR-2001; 2001US-0285030P.
PR 06-FEB-2002; 2002US-0355514P.
XX
PA (SCRI) SCRIP9 RES INST.
XX
PI Schultz P, Wang L, Anderson JC, Chin JWK, Liu DR, Magliery TJ;
PI Meggers EL, Mehl RA, Pasternak M, Santoro SW, Zhang Z;
DR WPI, 2003-120430/11.
XX
XX
PT Composition useful for producing protein comprising unnatural amino acid,
PT has translation system comprising orthogonal tRNA and orthogonal
PT aminoacyl tRNA synthetase.
XX
PS Example 4; Page 127-128; 188pp; English.
XX
XX
CC The invention relates to a novel composition comprising a translation
CC system comprising an orthogonal tRNA (O-cRNA) and an orthogonal aminoacyl
CC tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O
CC -cRNA with at least one unnatural amino acid in the translation system
CC and the O-cRNA recognises at least one selector codon. A composition of
CC the invention is useful for producing at least one protein comprising at
CC least one unnatural amino acid. The protein is the Asp127AG mutant of
CC chloramphenicol acetyltransferase (CAT), the Tyr163TAG mutant of mouse
CC dihydrofolate reductase (DHFR), or a Tyr163TAG mutant of mouse
CC dihydrofolate reductase comprising CHOCH His6tag. The unnatural amino
CC acid is provided exogenously. The translation system is a cell and the
CC unnatural amino acid is biosynthesised by the cell. The present sequence
CC represents a plasmid containing the individual genes papABC that encode

```



CC the enzymes used to carry out the conversion of chorismate to the  
XX unnatural amino acid p-aminophenylalanine (pAF)  
SQ Sequence 3305 BP; 445 A; 1327 C; 1121 G; 412 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	4.38e-213	Length:	3305
Score:	3127.50	Matches:	618
Percent Similarity:	91.40%	Conservative:	9
Best Local Similarity:	90.09%	Mismatches:	42
Query Match:	87.53%	Indels:	17
DB:	8	Gaps:	3

US-10-089-514-2 (1-686) x ABZ69799 (1-3305)

```
QY 1 MetArgThrIleuLeuIleAspAsnTYrAspSerPheThrHisAsnLeuPheGlnTYrIle 20
DB 1 ATGGGACGCTTGTGATCGACCACTACGACTCGTTCCACGAACTCTTCCAGTACATC 60
QY 21 GYGLVALAthrGlyGlnProValValProAsnAspAlaAspTyrSerArgLeu 40
DB 61 GGGAGGCGCACCGGGACGCCCC--GTCTGCCCCACGACCGCCGACTGTCGGCTG 117
QY 41 ProValGluAspPheAspAlaIleValSerProGlyProGlySerProAspArgGlu 60
DB 118 CCCCTCGAGGACTTCGACCGCGATCGTGTCTCCCGGGCCCGGACGCGCCGAG 177
QY 61 ArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValIleGlyVal 80
DB 178 CGGAGCTTCGGAGTACGCCCCCGGGCGATCACCGACGCGCTCGCCGTCCTCGGCTC 237
QY 81 CysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaProGlu 100
DB 238 TGGCTGGGCGCACGAGGCGATCGCCGAG-----CTCTCGGCGGAA 276
QY 101 ProMetHisGlyValArgValSerGluValArgHisThrGlyGluAspValPheArgGlyLeu 120
DB 277 CCCATGACGCGCGGGCTCCGAGGTGCGGACACCGGCGAGGACGTCTTCGCGGGCTC 336
QY 121 ProSerProPheThrAlaValArgTYrHisSerLeuAlaAlaThrAspLeuProAspGlu 140
DB 337 CCTCCCGCTTCCACGCGCGCTACCTCCCTCGCCGACCGACCGACTCCCGGACGAG 396
QY 141 LeuGluProLeuAlaTyrSerAspAspGlyValValMetGlyLeuArgHisArgGlyLeu 160
DB 397 CTGGAACCTCTCCCTGAGCGACGACGAGGCTGTGCTAGGGCTTGGCGACCCCGAGAG 456
QY 161 ProLeuTyrGlyValGlnPheHisProGlySerIleGlySerAspPheGlyArgGlyIle 180
DB 457 CGGCTATGGGCGCTCCAGTCCACCGGAGTCCATCGGACGCACTTCGCGCGGAGATC 516
QY 181 MetAlaAsnPheArgAspLeuAlaLeuAlaHisIleArgAlaArgArgHisGlyAlaAsp 200
DB 517 ATGGCAAACTTCGCGCACTCGCCCTCGCCACACCGGCGACGTGCGGACGCGGCGGAC 576
QY 201 SerProTyrGluLeuHisValArgArgValAspValLeuProAspAlaGlnGluValArg 220
DB 577 TGGGCTACCAACTCCACGTCGCGCGGTGACGTGTGCGGACCGCGAAGGATAGCG 636
QY 221 ArgGlyCysLeuProGlyGlnGlyThrThrPheTyrLeuAspSerSerSerValIleGlu 240
DB 637 CGGCGTGGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 696
QY 241 GlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGlnTYrLeuThr 260
DB 697 GGGGCTCGCGCTTCTCTTCTTCGCGGAGACCGGCGGCGGCGGCGGCGGCGGCGG 756
QY 261 TyrArgValAlaAspGlyValValSerValArgGlySerAspGlyThrThrThrArgThr 280
DB 757 TACCGGATCGCGACGCGGTGTCTCGGTCGCGGCTCCGACGCGACCAAGACCGGAGAC 816
QY 281 ArgArgProPhePheAsnTYrLeuGlnGluGlnLeuGlnArgArgValProValAla 300
DB 1870 GCGGACGCTTTCAGTACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1929
```

```
DB 817 GCGGACGCTTTCAGTACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 876
QY 301 ProGluLeuProPheGlnPheAsnLeuGlyTYrValGlyTYrLeuGlyTYrGlyLeuIle 320
DB 877 CCCGACCTGCTTTCAGTACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 936
QY 321 AlaGlnThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe 340
DB 937 GCGGAGACCAACCGGCGGACCGCGGAGTACCGGCGGCGGCGGCGGCGGCGGCGGCGG 996
QY 341 AlaAspArgAlaIleAlaLeuAspHisGlnGlyGlyCysGlyTYrLeuLeuAlaLeuAsp 360
DB 997 GCGGACGCGGCGCTCCGCTCGACCAACGAGAGAGGCTGCTACTGCTGCGGCTTCGAC 1056
QY 361 ArgArgGlyHisAspAspGlyValAlaArgAlaTyrPheLeuArgGlnThrAlaGlnThrLeuThr 380
DB 1057 CGCGGAGCGCACACACGCGCGCGCTGCTGCGGAGACGCGCGGAGACCTTCAC 1116
QY 381 GlyLeuAlaValArgAlaProAlaGlnProThrProAlaMetValPheGlyIleProGlu 400
DB 1117 GCGCTGCGGCTCGCGCTCGGCGGCGGCGGAGCCCGCCATGCTTTCGGGCTCCCGAG 1176
QY 401 AlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHisAspIleAspIleAspAlaTYrLeuIle 420
DB 1177 GCGGCGGCGGCTTCGCGCGCGCTCGCTCGGAGCGGCGGCGGCGGCGGCGGCGGCGG 1227
QY 421 ArgIleAspGluCysLeuIleGlnArgHisGlnGlySerTYrGlnIleCysLeuThr 440
DB 1228 -----TCGGGCTCGCGACCGGAGAGTGTACAGATCTGCTGAC 1269
QY 441 AsnMetValThrAlaProThrGlnAlaThrAlaLeuProLeuTyrSerAlaLeuArgAla 460
DB 1270 AACATGTCACCGCGCGGACCGAGCGGAGCGGCTCCGCTACTCCGCGCTCGCCGCG 1329
QY 461 IleSerProValProTyrGlyAlaLeuLeuGlnPheProGluLeuSerValIleLeuSerAla 480
DB 1330 ATCAGCCCGCTCGCGCTCGCGCTCGCTCGAGTCCCGAGCTCGGAGCTCACGCGC 1389
QY 481 SerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGlnSerIlePheIle 500
DB 1390 TCGCCGAGCGGCTCTCAAGTCCGCGGCGGAGCGGCGGCTCGAGTCCAGGCCATCAAG 1449
QY 501 GlyThrArgProArgGlyGlyIleThrAlaGlnGluIleAspGluArgLeuArgAlaAspLeuAla 520
DB 1450 GGGACCGCGCGCGGCGGCGGCGGCGGCGGAGAGAGAGAGGCTCCGCGCGGACTGAGC 1509
QY 521 GlyArgGluIleAspArgAlaGlnLeuMetIleValAspLeuValArgAsnAspLeu 540
DB 1510 GCGCGGAGAGAGAGCGGCGGCGGAGACGTGATGTGTGACTGCTCGGCAACGACCTC 1569
QY 541 AsnSerValCysAlaIleGlySerValHisValProArgLeuPheGlnValGlnThrTYr 560
DB 1570 AACAGGCTGCGCGATCGGCTCCGTCACAGTCCCGGCTCTTCAGAGTGGAGAGACTTC 1629
QY 561 AlaProValHisGlnLeuValSerThrIleArgGlyArgLeuAspProGlyThrSerThr 580
DB 1630 GCGCGCTGACCACTGATGTGTGACATCGGAGAGCGGCTGCGGCGGCGGCGGCGGCGG 1689
QY 581 AlaAlaCysValArgAlaAlaPheProGlyGlyIleSerMetThrGlyAlaProIleIleArg 600
DB 1690 GCGCGCTGATGCGCGCGCTTCCCGCGGCTCATACCGGCGGCGGCGGCGGCGGCGG 1749
QY 601 ThrMetGlnIleIleAspArgLeuGlnGlyProArgGlyValTYrSerGlyAlaLeu 620
DB 1750 CCCATGAGATATCAACGCTCGAGAGAGAGCGGCGGCGGCTTCACCGGCGGCGCTC 1809
QY 621 GlyTyrPheAlaLeuSerGlyValAlaAspLeuSerIleValIleArgThrIleValIleu 640
DB 1810 GGAATGTTGCGCTCAGCGGCGCGCGGCGGAGCTTCAAGATGTATCCGACCAATCGGCTG 1869
QY 641 AlaAspGlyGlnAlaGlnPheGlyValGlyIleValSerLeuSerAspGlnGlu 660
DB 1870 GCGGACGCGCGGCGGAGTTCGCGGTGCGGCGGCGGAGTGTCTCTCCGACCGAGAG 1929
```



Oy	661	GUUGUUPhETHrGluThVtValIlyAlaArgAlaMetValTmRAleuAspGlySer	680
Dd	1930	GAGGGATTCCAGCAACCGCTGTGCATGAAGCCGCCCATGTGCACC3CCTTGACGGCAGC	1989
Oy	681	AAValAIAGLyAlaArg	686
Dd	1990	GCAGTGGCGGGCCCCGA	2007
RESULT 4			
ID	ABZ69798	standard; DNA; 12391 BP.	
XX	AC	ABZ69798;	
DT	08-APR-2003	(first entry)	
XX	DE	Plasmid plasc-papabc.	
KM	DHFR; translation; orthogonal tRNA; O-tRNA; O-RS; CAT; TYRS;		
KM	orthogonal aminoacyl tRNA synthetase; unnatural amino acid;		
KM	chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;		
KM	biosynthesis; p-aminophenylalanine; pAF.		
XX	Synthetic.		
OS	WO200285923-AZ.		
PN	31-OCT-2002.		
PD	19-APR-2002; 2002MO-US012465.		
PF	19-APR-2001; 2001US-0286503P.		
PR	06-FEB-2002; 2002US-0355514P.		
XX	(SCRI ) SCRIPPS RES INST.		
PA	Schultz P, Wang L, Anderson JC, Chin JWK, Liu DR, Magliery TJ,		
PI	Meggers EL, Mehl RA, Pasternak M, Santoro SW, Zhang Z;		
DR	WPI; 2003-120430/11.		
XX	Composition useful for producing protein comprising unnatural amino acid,		
PT	has translation system comprising orthogonal tRNA and orthogonal		
PT	aminoacyl tRNA synthetase.		
XX	Example 4; Page 124-127; 188pp; English.		
PS	The invention relates to a novel composition comprising a translation		
XX	system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl		
CC	tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O		
CC	-tRNA with at least one unnatural amino acid in the translation system		
CC	and the O-tRNA recognises at least one selector codon. A composition of		
CC	the invention is useful for producing at least one protein comprising at		
CC	least one unnatural amino acid. The protein is the Asp12TAG mutant of		
CC	chloramphenicol acetyltransferase (CAT), the Tyr163TAG mutant of mouse		
CC	dihydrofolate reductase (DHFR), or a Tyr163TAG mutant of mouse		
CC	dihydrofolate reductase comprising CHOHH His6tag. The unnatural amino		
CC	acid is provided exogenously. The translation system is a cell and the		
CC	unnatural amino acid is biosynthesised by the cell. The present sequence		
CC	represents a plasmid for use in the biosynthesis of p-aminophenylalanine		
XX	(pAF) in vivo		
XX	Sequence 12391 BP; 2830 A; 3588 C; 3346 G; 2627 T; 0 U; 0 Other;		
Alignment Scores:			
Pred. No.:	2,06e-212	Length:	12391
Score:	3127.50	Matches:	618
Percent Similarity:	91.40%	Conservative:	9
Best Local Similarity:	90.09%	Mismatches:	42
Query Match:	87.53%	Indels:	17
DB:	8	Gaps:	3

	US-10-089-514-2 (1-686) x ABZ69798 (1-12391)
OY	1 MetArgThrLeuLeuIleAspAsnTyrAspSerPheThrHisAsmLeuPheGlnTyrIle 20
Db	24 ATGGCAAGCTTCGTGATGCACAACTAGACTCGTTCCACCAGAAGCTGTTCCAGTAATC 83
OY	21 GltgltAlaThrGlyGlnProProValValValProAsnAraAlaAspTrpSerArgLeu 40
Db	84 GGCGAGGCGCACCGGAGCAGCCCC--GTCGCGCCAACGACGCCGATGGTGCGGCTG 140
OY	41 ProValGluAspPheAspAlaIleValValSerProGlyLysPheProAspArgGlu 60
Db	141 CCCCTCGAGACTTCGACGCAGTGTGTGTCCCGGAGCCCCGAGCCCCCGAGCGGGA 200
OY	61 ArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeuGlyVal 80
Db	201 CGGAGCTTCGGAGTTCAGCCGCGCGGAGCATCACGACAGCGGCTGCGCTCTCGGCGTC 260
OY	81 CysLeuGlyHisGlnGlyTyrAlaGlnLeuPheGlyGlyThrValGlyLeuAlaProGlu 100
Db	261 TGCCCTGGCGCACCGAGGCATGCCAG-----CTCTCGCGGGA 299
OY	101 ProMetHisGlyArgValSerGlnValAlaGhisThrGlyGlyAsnAraPheArgGlyLeu 120
Db	300 CCATATGACGGCGCGGCTCTCGAAGTGGGAGAACCGCGGAGAGAGCTTCTCGGAGCTC 359
OY	121 ProSerProPheThrAlaValArgTyrHisSerLeuAlaAlaThrAspLeuProAspGlu 140
Db	360 CCTGTGCGGTTACCGCGCGTCCGCTACCACTCCCTGGCGCGCACCGACTCCCGACAG 419
OY	141 LeuGluProLeuAlaTrpSerAspAspGlyValValMetGlyLeuArgHisArgGluVal 160
Db	420 CTCGAACCCCCTGGCTGGAGCGAGCAGCGGCTGTCATGGCTCGGCGCACCGGAGAG 479
OY	161 ProLeuTrpGlyValGlnPheHisProGlnSerIleGlySerAspPheGlyArgGlnIle 180
Db	480 CCGCTGATGGCGGTCCAGTTCACACCGAGTCCATGGCACGACTTCGGCGGAGAGTTC 539
OY	181 MetaIaenPheaArgAspLeuAlaLeuAlaHisHisIstArgAlaArgArgHisGlyValAsp 200
Db	540 ATGGCGAACTTCGGGAGACTGCGCTCGGCCAACACCGGSGACGTCCGAGAGCGGCGAC 599
OY	201 SerProTyrGluLeuHisValArgArgValAspValLeuProAspAlaGlnValValArg 220
Db	600 TGGGCGTACGAATCCACGTGCGCGCGTCSAGCTGCGCGGACCGGAGAGGTATGCG 659
OY	221 ArgGlyCysLeuProGlyGlnGlyThrThrPheThrLeuAspSerSerValLeuGlu 240
Db	660 CGCGCTGCTGCCCGCGCGAGGGGCGCACGTTCTGGCTGGACACAGTCTCGTCTCGAA 719
OY	241 GlyIaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGlnTyrLeuThr 260
Db	720 GAGCGCTCGCGTCTCTCTCTCTGCGGACGACCGCGGCGCGCTCGCGGAGTACTCAC 779
OY	261 TyrArgValAlaAspGlyValValSerValArgGlySerAspGlyThrThrArgThr 280
Db	780 TACCGCGTCCGCGACGGGCTGTCTCGTCCGCGGCTCCGACGGGACCAACGACCGGGAC 839
OY	281 ArgArgProPhePheAsnTyrLeuGlnGlnLeuGluArgArgValProValAla 300
Db	840 GCGGCGACCCCTCTTCAGCTTACTGAGGAGCAGCTCGAACCGCGGCGGCTCCGCTCGC 899
OY	301 ProGluLeuProPheGluPheAsnLeuGlyTyrTyrAlaGlyTyrLeuGlyTyrGlnLeuVal 320
Db	900 CCCGACCTGCGCTTCGAGTTCACTCGGCTCACTCGGCTACCTCGGTACGACTGAAG 959
OY	321 AlaGluThrThrArgIyaSPProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe 340
Db	960 GCGGAGACCAACCGGCGACCCCGGAGTACCGGCCCCCGACCCCGACGCCGCTCTCTTC 1019
OY	341 AlaAspArgAlaIleAlaLeuAspHisGlnGlnGlyCysCysTyrLeuLeuAlaLeuAsp 360
Db	1020 GCCACACGCGGCATTCGCTCTGACCAACAAGGAGGCTGCTCTACCTGCTGGCTCGAC 1079



QY	361	AAGAAGGAGTAAASPAASGGLVAlaArgAlaATPLeuAArgGluThThAlaGluThThLeuThr	380
Db	1080	CGCCGGGGCCAGACGACGGCCCCCGCGCTGCGTCCGGAGACGGCCGAGACCTTCAAC	1139
QY	381	GIYLeuAlaValArgAlaProAlaGluProThrProAlaMetValPheGlyTLeuProGlu	400
Db	1140	GGCGTGGCGGCTCGCGGTCCGGCCGAGGGCCGACCCCGGCATGCTTTGGGGTCCCGAG	1199
QY	401	AlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHISaAspLYsaSPAlaTyrLeuLYs	420
Db	1200	GGGGGGCGCGGCTTGGGCCCCCTGGCTCGCGCACCCGACGACAGAGAGCC-----	1250
QY	421	ArgTILeaSPGluCysLeuLYseGluTLeaArgaNGlyuSerTyrGluTILeCysLeuThr	440
Db	1251	-----TCGGCGCTCGGACAGCGCGAGTGTGAAGATCTGCTTGACC	1292
QY	441	AsnMetValThThAlaProThrGluAlaThThAlaLeuProLeuTyrSerAlaLeuAArgAla	460
Db	1293	AACATGTGTACCGCGCCGACGAGCGACGGCCGCGGCTCTACTCCGGCTGGCGCC	1352
QY	461	ILeserProValProTyrGlyValaLeuLeuGluPheProGluLeuSerValLeuSerAla	480
Db	1353	ATCAGACCCCGTCCCGTCCGGCGCCCTGTCTCGATTCCCGAGCTGTCTGATCAGCGCC	1412
QY	481	SerProGluAArgPheLeuThrTILeGlyAlaAspGlyGlyValaGluSerLYsProILeLYs	500
Db	1413	TCGCGCGAGCGGTTCTCTACATCGAGCGCCCGACGGCGCGCTGTGATCCAAAGCCATCAAG	1472
QY	501	GIYThThAArgProAArgGlyTylThThAlaGluGluAspGluAArgLeuAArgAlaAspLeuAla	520
Db	1473	GGGAGCCCGCCCCCGGGGGCGACCGCGGAGGAGGAGCGAGGCTCCGCGCCGACTGGCC	1532
QY	521	GIYAArgGluLYsaSPAlaArgAlaGluAsnLeuMetILeValaSPLeuValaArgAsaSPLeu	540
Db	1533	GGCGCGGAGAGGACCGGGCGGAGAACTGTATGTCTGTCTGACCTGTCTCGGACAGACCTC	1592
QY	541	AsnSerValCysAlaTILeGlySerValHISValProAArgLeuPheGluValaGluThThTyr	560
Db	1553	AACAGCGTCTGGCGATCGGCTCGCTCCATCGACCTGGCCCCGCTCTTGAGAGTGGAGACCTC	1652
QY	561	AlaProValHISGluLeuValSerThrTILeaArgGlyAArgLeuAArgProGlyThThSerThr	580
Db	1653	GGCGCCGCGACCAAGCTGTGTGTGACCATCCGAGGAGCGGTGCGGCGCGGACGACGACC	1712
QY	581	AlaAlaCysValaArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLYsLYsArg	600
Db	1713	GGCGGCTCGTACGCGCGCGCTTCCCGCGGCTCTCAAGACGGCGCGGCCCAAGACGGA	1772
QY	601	ThreMetGluTILeIeaSPArgLeuGluGluGlyProAArgGlyValaTyrSerGlyAlaLeu	620
Db	1773	CCCATGGAGATCATGACCGCTCTGGAGGAAGAGCCCCCGAGGCGCTTACCGGGGGCGCTC	1832
QY	621	GIYTPPheAlaLeuSerGlyAlaAlaAspLeuSerILeValaTILeaArgThrILeValLeu	640
Db	1833	GGATGTGTTCGCCCTTCAGCGCGCGCGGACCTCAGACATCGTCAATCCGACCATGTGCTG	1892
QY	641	AlaAspGlyGluAlaGluPheGlyValaGlyGlyAlaILeValSerLeuSerAspGluGlu	660
Db	1893	GCCAGCGCGCGGCGGAGTGTGGGCTCGCGCGGCGGATCGTGTCTCTCCGACAGAGAG	1952
QY	661	GluGluPheThrGluThThValaLYsValaLYsAlaArgAlaMetValThThAlaLeuAspGlySer	680
Db	1953	GAGAGATTCAAGCAGACCGTGTGTCAAGCCCGCCCATGTGTACCGCCCTCGAGCGAGAGC	2012
QY	681	AlaValaAlaGlyAlaArg	686
Db	2013	GCAGTGGCGGCGGACGGA	2030

RESULT 5

AAH6076

ID AAH6076 standard; DNA; 1860 BP.

XX

AAH6076 standard; DNA; 1860 BP.

AC	AAB66076;
XX	
DT	26-SEP-2001 (first entry)
XX	
DE	C glutamicum coding sequence fragment SEQ ID NO: 1111.
XX	
KW	Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX	
XX	organic acid synthesis; ds.
OS	Corynebacterium glutamicum.
XX	
PN	EPI108790-A2.
PD	20-JUN-2001.
XX	
PF	18-DEC-2000; 2000EP-00127688.
XX	
PR	16-DEC-1999; 99JP-00377484.
PR	07-APR-2000; 2000JP-00159162.
PR	03-AUG-2000; 2000UP-00280988.
XX	
PA	(KYOW ) KYOWA HAKKO KOGYO KK.
XX	
PI	Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI	Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX	
DR	WPI; 2001-376931/40.
XX	P-PADB; AAG90857.
PT	
PT	Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT	mutation point of a gene, measuring expression of a gene, analyzing
PS	expression profile or pattern of a gene and identifying homologous gene.
XX	
PS	Claim 8; SEQ ID NO 1111; 246pp + Sequence Listing; English.
XX	
CC	The present invention provides a number of nucleotide and protein
CC	sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
CC	are useful for identifying the mutation point of a gene derived from a
CC	mutant of coryneform bacterium, measuring expression amount and analysing
CC	the expression profile or expression pattern of a gene derived from
CC	Coryneform bacterium, and identifying a homologue of a gene derived from
CC	coryneform bacterium. Coryneform bacteria are useful for producing amino
CC	acids, nucleic acids, vitamins, saccharides and organic acids,
CC	particularly L-lysine. The present sequence is a nucleic acid described
CC	in the exemplification of the invention. Note: The sequence data for this
CC	patent did not form part of the printed specification, but was obtained
CC	in electronic format directly from the European Patent Office
XX	
SQ	Sequence 1860 BP; 394 A; 478 C; 516 G; 472 T; 0 U; 0 Other;
	Alignment Scores:
	Prod. No.: 1.61e-96 Length: 1860
	Score: 1487.00 Matches: 328
	Percent Similarity: 60.06% Conservative: 78
	Best Local Similarity: 48.52% Mismatches: 204
	Query Match: 41.62% Indels: 66
	DB: 5 Gaps: 13
US-10-089-514-2 (1-686) x AAB66076 (1-1860)	:
OY	1 MetATGThTleuleuilealeapAntYrApSeSPerhethrhiasnleuPhegIntyrille 20
Db	1 ATCGCGGTATTATTTATGTATTAATTAATTCCTTCAAGCCTTAATCTGCACCATATGTG 60
OY	21 GLYGluAlathnGlyGlnProProValValValP.ProabnapAlaepTPserArgLeu 40
Db	61 GAAGAGGTTACGGGTACGACACTGTGTGGTGCCCTTAATGATCAAGAAMTAGATGAGATG 120
OY	41 ProValGluaSpPheaspAlaIleValValSerProGlylProGlylSerProAspArgLlu 60
Db	121 CTT-----TTGACGCCGCTACATCTCTCACCTGACCAGGCGACAGCGCGTTGCCG 171
OY	61 ArgaspPheGlylIeserArgAlalalleThrasPserGlyleuProValleuGlyVal 80



Db 172 GCTGATTTTGGATCTGTGCAGCGCTCATTTAGCGTGCACGCGTTCCATTTTGGGTGTG 231  
OY CysIeuGIyH:sgIuGIyIleAlaGIuLeuPheGIyIyThrValGIyLeuAlaProGIu 100  
Db 232 TGTTTAGCGCCACACAGGGCATTTGCGTGTGCGCTATGCGCGGTGATGTTGATTGGCGCCAGG 291  
OY 101 PromethIsgIyAArgValSerGIuValAArgHsthIRGIyGIuAspValPheArgGIyLeu 120  
Db 292 CCGGTCCACGGTGAGGTTTGGAGATTCACCAATGATGGTTCAGGTTATTGGACAGCAATC 351  
OY 121 ProSerProPheThAlaValAArgYHrhIserLeuAlaAlaThAspLeuProAspGIu 140  
Db 352 CCGTGAACGTTTGAGGCGGTGTGTTATCACTGATGAGTGGCCAAACCGCTTGCCGAGTCA 411  
OY 141 LeuGIuProLeuAlaIyTrpSerAspAspGIyValValMetGIyLeuAArgHIsArgGIuIys 160  
Db 412 TTGAAGAGCTACAGCTACACGCGATGATGTTGATCATGGATGGACATGAAGTGCTT 471  
OY 161 ProIeuTrpGIyValGIuPheHsthIerProGIuSerIleGIySerAspPheGIyAArgGIuIle 180  
Db 472 CCGCAGTGGGGGTGCAATTTTCATCCGAATATATTGGTGCAATTTGGCGCATCAGATC 531  
OY 181 MetaIaenPheAArgAspLeuAlaLeuAlaHsthIsArgValAArgArgHIsGIyAlaAsp 200  
Db 532 ATTAAGATCTCTTAATTTAGCGCGCACATATGCC----- 567  
OY 201 SerProYrGIuLeuHIsValAArgAArgValAspValLeuProAspAlaGIuGIuValAArg 220  
Db 568 -----TGGCAACTCAGCGAGAAACATATTCGCGTCAGCGTTGATTCAAGACGGTTTTT 621  
OY 221 ArgGIyCysLeuProGIyGIuGIyIyThrThrPheTrpLeuAspSerSerSerValLeuGIu 240  
Db 622 GAACCATTTCTTGGCCATTTCTCCCATGCTTTTGGCTCGATGAT----- 666  
OY 241 GIyAlaSerAArgPheSerPheLeuGIyAspAspAArgGIyProLeuAlaGIuIyLeuThr 260  
Db 667 ---GCCCAAGAAACAGCTATCTGTGTATGTCACAGCGGCTCTTCGACGCAAAAAACC 723  
OY 261 TyrAArgValAlaAspGIyValValSerValAArgIySerAspGIyThrThrAArgThr 280  
Db 724 CATATATGCGGGAGAGG----- 741  
OY 281 AArgAArgProPhePheAsnIyLeuGIuGIuGIuLeuGIuAArgAArgValAProValAla 300  
Db 742 -----GATTTCTTCACTCGCTTAAGAGATCTCGCGCCAAC-----TCAAGTTGCG 789  
OY 301 ProGIuLeuProPheGIuPheAsnLeuGIyYrValGIyTrpLeuGIyYrGIuLeuIys 320  
Db 790 CCGCGTCAA-----GGTTTCGTGTGGCTGGGGTGTGTACGTGGTTATGAGCTTAA 843  
OY 321 AlaGIuThrThrGIyAspProAlaHsthAsrSerProHIsProAspAlaAlaPheLeuPhe 340  
Db 844 GCGGAAGCTGGCGCACCGGGCTGCGCACATTCGAGTCTTCGCGATGCCACCTCATTTT 903  
OY 341 AlaAspAArgAlaIleAlaLeuAspHsthIsgIuGIyCysCysTrpLeuLeuAlaLeuAsp 360  
Db 904 GCCGATCCGCGCATCGCAGTGAATCGGATCAG-----GTTCCGTTCTGGCGTTG-- 954  
OY 361 AArgAArgGIyHIsAspAspGIyAlaAArgAlaTrpLeuAArgGIuThrAlaGIuThrLeuThr 380  
Db 955 -----GGGGAGCAGAGAC-----GAGTGGTTTGAAGAAACATCAAGAACGTGCAT 999  
OY 381 GIyLeu---AlaValAArgAlaProAlaGIuProThrProAlaMetValPheGIyIlePro 399  
Db 1000 AATCTTTCGCGCGCGCGGATACCTTCGCTC----- 1029  
OY 400 GIuAlaAlaAlaGIyPheGIyProLeuAla---AArgAlaArgHIsAspIysAspAlaIyTr 418  
Db 1030 -----GGAACCTCGCTTTCGAGGTTCCAGATTCGAAGATTCGAAGATTCAGAT 1071  
OY 419 LeuIysAArgIleAspGIuCysLeuIysGIuIleAArgAsnGIyIySerTrpGIuIleCys 438

Db	1072	TTGCACAAAAATTCGACAGCGCCACGAGCTGATTACTCGCGCGCATGTGTAAGAAATCTGC	1131
QY	439	LeuThrIshmetValIThrAlaProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeu	458
Db	1132	CTGACCCACAAAACCTTCAGGCGCACCACTGATGTGGCCCTTCGCGCTCATCTAGCACTG	1191
QY	459	ArgAlaIleSerProValProTyrGlyValAlaLeuLeuGluPheProGluLeuSerValLeu	478
Db	1192	CGTGGGGCCAAATCCACCGCATATGAGCGCATCTTCACGTGGGGGAACTCTAATTTTG	1251
QY	479	SerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyValGluSerLysPro	498
Db	1252	AGTTCCTCGCGGAGCGGTTATCATCACTTGAATTCGGAGGGATATGTGATCAAAAGCC	1311
QY	499	IleIysGlyThrArgProArgGlyValGlyThrAlaGluGluAspGluArgLeuArgAlaAsp	518
Db	1312	ATTAAAGGACACAGCGCGCGTGGCGCAACGCCAGCAAGACCAAGAAATCATGTGCTGAG	1371
QY	519	LeuAlaGlyArgGlyIlyAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsn	538
Db	1372	CTGGCGCATTAATCTTAAAGATCGGCAAGAAATCTGATGATCGTGATTTGGTCCGCAAC	1431
QY	539	AspLeuAsnSerValCysAlaIleGlySerValIshValProArgLeuPheGluValGlu	558
Db	1432	GACTTAGCGCGCGCGCGCTTGCCACACAGATTTAAACATCCAAAGCTTTTGACGTGCA	1491
QY	559	ThrTyrAlaProValIshGluLeuValSerThrIleArgGlyArgLeuArgProGlyThr	578
Db	1492	ACCTTAGCGCACAGTCCACCAACTGTGTACAGCACCGTCTGTGACAGATTGGGGCCA---	1548
QY	579	SerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyValProLys	598
Db	1549	AGTCCGATTGAGTGCAGCGCGCGCAACTTCCCGGTGTGTGAGTCTGGGCCCAAG	1608
QY	599	LysArgThrMetGluIleIleAspArgLeuGluGlyProArgGlyValTyrSerGly	618
Db	1609	CTGCGGCACATGAGATCATGATGAGCTGGAGGACAGCTCTCCCGATTATTCTCAGGT	1668
QY	619	AlaLeuGlyTrpPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIle	638
Db	1669	GGCTTGGGATATTTTCCCTCGACGGGCGAGTTGATCTCTCATGATGATCAAGACTCTC	1728
QY	639	ValLeuAlaAspGlyGlnAlaGluPheGlyValGlyValAlaIleValSerLeuSerAsp	658
Db	1729	GTCATCCAGAACATCATCTGAGACTAGCGAGTGGCGGTGACATTCTTGCTCTGTGAT	1788
QY	659	GlnGluGluGluPheThrGluThrValValLysAlaArgAlaMetVal	674
Db	1789	CCGAGGCTGAGTGGAGAGAAATCCGGCTTAATCAAGCGCTTGCTG	1836
RESULT 6			
AAAF72103			
ID	AAAF72103	standard; DNA; 1983 BP.	
AC	AAAF72103;		
XX			
DT	30-APR-2001	(first entry)	
XX			
DE	Corynebacterium glutamicum MP protein nucleotide sequence SRQ ID NO:701.		
XX			
KM	Corynebacterium glutamicum; metabolic pathway protein; MP protein;		
KW	fine chemical production; microorganism; organic acid; nucleoside;		
KW	nonproteinoic amino acid; purine base; pyrimidine base; nucleotide;		
KW	lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;		
KW	carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.		
XX			
OS	Corynebacterium glutamicum.		
XX			
XX	MO200100843-A2.		
PN	04-JAN-2001.		
PD			
XX			
XX	23-JUN-2000; 2000OWO-IB000923.		



XX	25-JUN-1999;	99US-0141031P.
PR	01-JUL-1999;	99DE-01030476.
PR	02-JUL-1999;	99US-0142101P.
PR	08-JUL-1999;	99DE-01031415.
PR	08-JUL-1999;	99DE-01031418.
PR	08-JUL-1999;	99DE-01031419.
PR	08-JUL-1999;	99DE-01031420.
PR	08-JUL-1999;	99DE-01031424.
PR	08-JUL-1999;	99DE-01031434.
PR	08-JUL-1999;	99DE-01031435.
PR	08-JUL-1999;	99DE-01031478.
PR	08-JUL-1999;	99DE-01031510.
PR	08-JUL-1999;	99DE-01031541.
PR	08-JUL-1999;	99DE-01031573.
PR	08-JUL-1999;	99DE-01031592.
PR	08-JUL-1999;	99DE-01031632.
PR	08-JUL-1999;	99DE-01031634.
PR	08-JUL-1999;	99DE-01031636.
PR	09-JUL-1999;	99DE-01032125.
PR	09-JUL-1999;	99DE-01032126.
PR	09-JUL-1999;	99DE-01032130.
PR	09-JUL-1999;	99DE-01032186.
PR	09-JUL-1999;	99DE-01032206.
PR	09-JUL-1999;	99DE-01032227.
PR	09-JUL-1999;	99DE-01032228.
PR	09-JUL-1999;	99DE-01032229.
PR	09-JUL-1999;	99DE-01032230.
PR	14-JUL-1999;	99DE-01032922.
PR	14-JUL-1999;	99DE-01032926.
PR	14-JUL-1999;	99DE-01032928.
PR	14-JUL-1999;	99DE-01033004.
PR	14-JUL-1999;	99DE-01033005.
PR	14-JUL-1999;	99US-0148613P.
PR	12-AUG-1999;	99DE-01040764.
PR	27-AUG-1999;	99DE-01040765.
PR	27-AUG-1999;	99DE-01040766.
PR	27-AUG-1999;	99DE-01040832.
PR	31-AUG-1999;	99DE-01041378.
PR	31-AUG-1999;	99DE-01041379.
PR	31-AUG-1999;	99DE-01041380.
PR	31-AUG-1999;	99DE-01041394.
PR	31-AUG-1999;	99DE-01041396.
PR	03-SEP-1999;	99DE-01042076.
PR	03-SEP-1999;	99DE-01042077.
PR	03-SEP-1999;	99DE-01042079.
PR	03-SEP-1999;	99DE-01042086.
PR	03-SEP-1999;	99DE-01042087.
PR	03-SEP-1999;	99DE-01042095.
PR	03-SEP-1999;	99DE-01042098.
PR	03-SEP-1999;	99DE-01042124.
PR	03-SEP-1999;	99DE-01042129.
PR	09-MAR-2000;	2000US-0187970P.
XX	(BADI ) BASF AG.	
PA		
XX	Pompejus M, Kroegeer B, Schroeder H, Zeidler O, Haberhauer G;	
XX	WPI, 2001-137957/14.	
DR	P-PSDB; AAB79984.	
XX		
PT	Nucleic acids from Corynebacterium glutamicum encoding metabolic pathway	
PT	proteins, useful for producing fine chemicals in microorganisms,	
PT	including organic acids, nonproteinogenic amino acids, and purine and	
PT	pyrimidine bases.	
XX		
PS	Claim 3; Page 1130-1133; 1737P; English.	

XX	AAF71753	0	encode the	Corynebacterium	glutamicum	metabolic	CC
CC	pathway (MP)	proteins	given in	AA079634	to	AA080211	The C. glutamicum
CC	nucleic acids	are useful	for the	production	of	fine chemicals	in
CC	microorganisms,	including	organic acids,	nonproteinogenic	amino acids,		
CC	purine and pyrimidine	bases, nucleosides,	nucleotides, lipids,	saturated			
CC	and unsaturated	fatty acids, diols,	carbohydrates,	aromatic compounds,			
CC	vitamins, cofactors,	polyketides	and enzymes				
XX	Sequence	1983 BP,	409 A;	500 C;	553 G;	521 T;	0 U;
SQ	0	Other;					
Alignment Scores:							
Pred. No.:	1,74e-96	Length:	1983				
Score:	1487.00	Matches:	378				
Percent Similarity:	60.06%	Conservative:	78				
Best Local Similarity:	48.52%	Mismatches:	204				
Query Match:	41.62%	Indels:	66				
DB:	4	Gaps:	13				
US-10-089-514-2 (1-686) x AAF72103 (1-1983)							
QY	1	MetArgThrLeuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGlnTyrIle	20				
DB	101	ATGGCGGTTTAAATTATGATATATATATATGTTCTTCACGTTTAATCCGCCACCTATG	160				
QY	21	GlyIuaIatThrGlyValProProValValValProAsnAspAlaAspTrpSerArgLeu	40				
DB	161	GAAGAGCTTACGGGCTCAGAGCACCTGTGGTGGCTTAAGATCAAGAAATAGATGAGATG	220				
QY	41	ProValGluAspPheAspAlaIleValValSerProGlyProGlySerProAspArgIu	60				
DB	221	CTT-----TTGACGCGCTCAATCCTCACTGGCGGAGCCAGCCCGCGCTTGGC	271				
QY	61	ArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeuGlyVal	80				
DB	272	GCTGATTTTGGATCTGTGCAGCGGCTATTGACGCGGACGCGGTTCCGATTTTGGGTGG	331				
QY	81	CysLeuGlyIleGlnGlyIleAlaGlnLeuPheGlyIleValGlyLeuAlaProGlu	100				
DB	332	TGTTTAAAGCCACAGGAGCATGGCGTTCATAGCGGATGATTTGGGCGCCAG	391				
QY	101	PromethIsgIArgValSerGlyValAlaArgHisThrGlyGluAspValPheArgGlyLeu	120				
DB	392	CGGCTCCACGGTGAAGCTTCCGACATCCACCTATGATGGTTACAGTTATTTCAGGCAATC	451				
QY	121	ProSerProPheThrAlaValArgTyrHisSerLeuAlaIleThrAspLeuProAspGlu	140				
DB	452	CCTGAACGTTTGAAGCGGCTGGTTATCACTGATGGTGGCAACCGCTTGGCGAGTCA	511				
QY	141	LeuGluProLeuAlaIleTrpSerAspArgValValMetGlyLeuArgHisArgGlyIys	160				
DB	512	TTGAAAGCTACAGCTACACGAGCATGATGGTTGATCATGGCATGGCACATGAAGTCTT	571				
QY	161	ProLeuTrpGlyValGlnPheHisProGluSerIleGlySerAspPheGlyArgGluIle	180				
DB	572	CCGCACTGGGAGTGTCAATTCATCCGAAATCATTTGGTGGCAATTCGCAATCCATCCATC	631				
QY	181	MetAlaAsnPheArgAspLeuAlaLeuAlaHisHisArgAlaArgArgHisGlyAlaAsp	200				
DB	632	ATTAAAGACTCTCTTAATTATGCGGCGCATATCC-----	667				
QY	201	SerProTyrGlyIleLeuHisValAlaArgArgValAlaSerValLeuProAspAlaGluIleValArg	220				
DB	668	-----TGGCAACTCAGCAGGAAACATATTCCGCTCACCGCTTGATTACAGACCGGTTT	721				
QY	221	ArgGlyCysLeuProGlyGluGlyThrThrPheTrpLeuAspSerSerValLeuGlu	240				
DB	722	GAACATCTCTTGGCCATCTCCACAGCTTTTGGCTCGATGAT-----	766				
QY	241	GlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGlyIleThr	260				
DB	767	---GCCCAAGAACACAGCTATCTTGGTATGCTCAGCGGCTCTCTGCAAGCAAAACCC	823				















```

QY 161 ProLeuTrpGlyValGlnPheHisProGluSerIleGlySerAspPheGlyArgGluIle 180
DB 587 CCGCAGTGGGGGTGTCATTTTCATCCGGAATCTATTGGTGACAAATTCGGCATCAGATC 646
QY 181 MetLalaPheAspPheLeuAlaLeuAlaHisIleArgAlaArgArgHisGlyAlaAsp 200
DB 647 ATTAAGAACTTCTTAATTAGCCGCGCAATATCGC----- 682
QY 201 SerProTyrGluLeuHisValArgArgValAlaLeuProAspAlaGluGluValArg 220
DB 683 -----TGGCACTTCACGAGAGAAAATTAATCCGCTCAGCGTTGATTCAGCAGCGGTTTTT 736
QY 221 ArgGlyCysLeuProGlyGluGlyThrThrPheTrpLeuAspSerSerValLeuGlu 240
DB 737 GAACATTTCTTGGCCCATCTCTCCATGCTTTTGGCTCGATGAT----- 781
QY 241 GlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTyrLeuThr 260
DB 782 ---GCCCAAGGAAACACGATCTTGTGTATGCCAGCGTCTCTCGACGCACAAAACC 838
QY 261 TyrArgValAlaAspGlyValValSerValArgGlySerAspGlyThrThrArgThr 280
DB 839 CATATGTCGGCGAGGGG----- 856
QY 281 ArgArgProPhePheAsnTyrLeuGluGluGlnLeuGluArgArgValProValAla 300
DB 857 -----GATTTCTTACCCTGGCTAAAGAGATCTCCGCCCAAC-----TCAGTTCCG 904
QY 301 ProGluLeuProPheGluPheAsnLeuGlyTyrValGlyTyrLeuGlyTyrGluLeuLys 320
DB 905 CCCGGTCAA-----GGTTTTCGCTTGGCTGGGTGGTTAAGTTAGCTTAA 958
QY 321 AlaGluThrThrArgLysAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe 340
DB 959 GCGGAACTGGCGCAGCGGCTCGCACACTTCGAGTCTTCGAGTGGCAGCACTATTTT 1018
QY 341 AlaAspArgAlaIleAlaLeuAspHisGlnGluGlyCysCysTyrLeuLeuAlaLeuAsp 360
DB 1019 GCCGATCGCCGATCGCATGCGAATCGGATCAG-----GTTCGTTGCTGGCGTTG-- 1069
QY 361 ArgArgGlyHisAspAspArgLysAlaArgAlaTyrPheArgGluThrAlaGluThrLeuThr 380
DB 1070 -----GGGAGCAGGAC-----GAGTCGTTTAAAGAAACCATCAAGAGCTGCAT 1114
QY 381 GlyLeu---AlaValAlaArgAlaProAlaGluProThrProAlaMetValPheGlyIlePro 399
DB 1115 AATCTTGTCCGCCCGCATACCTGCGTCC----- 1144
QY 400 GluAlaIleAlaGlyPheGlyProLeuAla---ArgAlaArgHisAspLysAspAlaTyr 418
DB 1145 -----GGACACCTCGCTTGGAGTTTGAGATTCCAAAGATGAGTAT 1186
QY 419 LeuLysArgIleAspGlyCysLeuLysGluIleArgAsnGlyGluSerTyrGluIleCys 438
DB 1187 CTGACAAATAATTCGACAGCCAGGAGCTGATTACTCGCGCGAATGTATGAATTCG 1246
QY 439 LeuThrAsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeu 458
DB 1247 CTGACCAAAACTTCAGGCGCACCATGATGCGCCCTGCGCTGCTCATAGCACTG 1306
QY 459 ArgAlaIleSerProValProTyrGlyAlaLeuLeuGluPheProGluLeuSerValLeu 478
DB 1307 CGTGGGCCAATCCACCCGATATGCTGCTTACGCTGGGGGATACCTCTATTTTG 1366
QY 479 SerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGluSerLysPro 498
DB 1367 AGTTCTCGCGCGAGGCTTCACTACCATTTGATTCGCGCAGGATATGCGAATCAAGCCC 1426
QY 499 IleLysGlyThrArgProArgGlyGlyTyrAlaGluGluAspGluArgLeuArgAlaAsp 518
DB 1427 ATTAAGGACACAGCGCGCTGGCGAAGACGCAAGAAACATCATTTGCTGAG 1486

```

```

QY 519 LeuAlaGlyArgGluLysAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsn 538
DB 1487 CTGGCAGGTAATCCTTAAGATCGTGCAAGAAAACCTTGATGATCTTGATTTGGTCCGCAAC 1546
QY 539 AspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGlu 558
DB 1547 GACTTAGCCCGCGCGCTTTCACCACACAGATTAAACATCAAGCTTTTCAGCTCGAA 1606
QY 559 ThrThrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyThr 578
DB 1607 ACCTAGCCCAAGTCCACCAACTTGTACGACCGCTCTCGACAGATTGGGCA---CGC 1663
QY 579 SerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLys 598
DB 1664 AGTCCAGTTAGAGTCGTGCGCGCAGCATTCGCCGTTGTTGATGATGCTGCTCCCAAAG 1723
QY 599 LysArgThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTyrSerGly 618
DB 1724 CTGCGCACCATGAGATCATCATGATGAGCTGGAGGAGCTCTCGCGTATTTACTCAGGT 1783
QY 619 AlaLeuGlyThrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIle 638
DB 1784 GCGTTGGATATTTTCCCTCGACGCGCAGATTATCTCCATGGTATCAGAACTCTC 1843
QY 639 ValLeuAlaAspGlyGlnAlaGluPheGlyValGlyValAlaIleValSerLeuSerAsp 658
DB 1844 GTATCCAGAAACAATCAGCTGAGTACGAGTGGGCGTGCACTTCTGCTCTGCTGAT 1903
QY 659 GlnGluGluGluPheThrGluThrValValLysAlaArgAlaMetVal 674
DB 1904 CCGGAGCGTGAAGTGGAGGAATCCGCGTTAATCACGGCGCTGTGCTG 1951

```

## RESULT 9

AAH68527 standard; DNA; 349980 BP.

AAH68527;

26-SBP-2001 (first entry)

C glutamicum coding sequence fragment SEQ ID NO: 7062.

KMW Corynebacterium; amino acid synthesis; vitamin; saccharide;

KMW organic acid synthesis; ds.

OS Corynebacterium glutamicum.

PN EP1108790-A2.

PD 20-JUN-2001.

PF 18-DEC-2000; 2000BP-00127688.

PR 16-DEC-1999; 99P-00377484.

PR 07-APR-2000; 2000JP-00159162.

PR 03-AUG-2000; 2000JP-00280988.

PA (KYOWA) KYOWA HAKKO KOGYO KK.

PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.

XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying

XX mutation point of a gene, measuring expression of a gene, analyzing

XX expression profile or pattern of a gene and identifying homologous gene.

XX Disclosure; SEQ ID NO 7062; 246bp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein

XX sequences from the Corynebacterium glutamicum Corynebacterium glutamicum. These

XX are useful for identifying the mutation point of a gene derived from a







QY	619	AlaLeuGlyTrpPheAlaLeuSerGlyValAlaAlaPheLeuSerIleValIleArgThrIle	638
Db	153689	GGCTTGGATATATTTTCCCTCCGACGGCCGACGTTGATCTCTCCATGGATCAGAACTCTC	153748
QY	639	ValLeuAlaAspGlyGlnAlaGluPheGlyValGlyGlyAlaIleValSerLeuSerAsp	658
Db	153749	GTCATCCAGAACATCACTGCGAGTACGAGAGTGGCGGCGGACCTTCTGCTGAT	153808
QY	659	GlnGluGluGluPheThrGluThrValIleValAlaArgAlaMetVal	674
Db	153809	CCGAGGCTGATGGTGGAGAAATCCGCTTAATCAGCGCCCTGCTG	153856
RESULT 10			
ACF67367_10			
Continuation (50 of 57) of ACF67367 from base 4900001 (Photobabidus luminescens nucleoti			
WP	Sequence split into 57 fragments		
WP	Fragment Name	Begin	End
WP	ACF67367_00	1	110000
WP	ACF67367_01	100001	210000
WP	ACF67367_02	200001	310000
WP	ACF67367_03	300001	410000
WP	ACF67367_04	400001	510000
WP	ACF67367_05	500001	610000
WP	ACF67367_06	600001	710000
WP	ACF67367_07	700001	810000
WP	ACF67367_08	800001	910000
WP	ACF67367_09	900001	1010000
WP	ACF67367_10	1000001	1110000
WP	ACF67367_11	1100001	1210000
WP	ACF67367_12	1200001	1310000
WP	ACF67367_13	1300001	1410000
WP	ACF67367_14	1400001	1510000
WP	ACF67367_15	1500001	1610000
WP	ACF67367_16	1600001	1710000
WP	ACF67367_17	1700001	1810000
WP	ACF67367_18	1800001	1910000
WP	ACF67367_19	1900001	2010000
WP	ACF67367_20	2000001	2110000
WP	ACF67367_21	2100001	2210000
WP	ACF67367_22	2200001	2310000
WP	ACF67367_23	2300001	2410000
WP	ACF67367_24	2400001	2510000
WP	ACF67367_25	2500001	2610000
WP	ACF67367_26	2600001	2710000
WP	ACF67367_27	2700001	2810000
WP	ACF67367_28	2800001	2910000
WP	ACF67367_29	2900001	3010000
WP	ACF67367_30	3000001	3110000
WP	ACF67367_31	3100001	3210000
WP	ACF67367_32	3200001	3310000
WP	ACF67367_33	3300001	3410000
WP	ACF67367_34	3400001	3510000
WP	ACF67367_35	3500001	3610000
WP	ACF67367_36	3600001	3710000
WP	ACF67367_37	3700001	3810000
WP	ACF67367_38	3800001	3910000
WP	ACF67367_39	3900001	4010000
WP	ACF67367_40	4000001	4110000
WP	ACF67367_41	4100001	4210000
WP	ACF67367_42	4200001	4310000
WP	ACF67367_43	4300001	4410000
WP	ACF67367_44	4400001	4510000
WP	ACF67367_45	4500001	4610000
WP	ACF67367_46	4600001	4710000
WP	ACF67367_47	4700001	4810000
WP	ACF67367_48	4800001	4910000
WP	ACF67367_49	4900001	5010000
WP	ACF67367_50	5000001	5110000
WP	ACF67367_51	5100001	5210000
WP	ACF67367_52	5200001	5310000
WP	ACF67367_53	5300001	5410000
WP	ACF67367_54	5400001	5510000
WP	ACF67367_55	5500001	5610000

[illegible]







QY	203	TyrGluLeuNH <sub>2</sub> SValAlaArgValAlaProAlaLeuProAspAlaGluGluValAlaArgGly	222
Db	14551	CTCAACCTGAATTATTCGTAATGTAAGAGATAGTCAATCTTGTAATCTTGTTATTCA	14611
QY	223	CysLeuProGlyGluGlyThrThrPheThrLeuAspSerSerValLeuGluGlyAla	242
Db	14611	CGTTATGGCAATGATAGCAATGCTTTTCGCTTATATGTAAATTCGTATAGCCAA	14671
QY	243	SerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTyrLeuThrTyrArg	262
Db	14671	GCCCCCTTCTCAATGATGGCGAGCGGTATGAACAAGGTCTGTCAACTGTAATACAGT	14731
QY	263	ValAlaAspGlyValAlaValSerValAlaGlySerAspGlyThrThrAlaGlyThrArg	282
Db	14731	GTACAGACTGAGAAATTTGCGTTTATTCGGTCTGTATGGCCAGAGAGATT--GTTCAAGGA	14787
QY	283	ProPhePheAsnTyrLeuGluGluGluLeuGluArgArgValProValAlaProGlu	302
Db	14788	GATTTTTTCTCGCGTATGTCACGTTCGTTGATGTGGTGGCAATTTGCAACCAAAATCG	14844
QY	303	LeuProPheGluPheAsnLeuGlyTyrValGlyTyrLeuGlyTyrGluLeuValGlu	322
Db	14848	ATGCCATTTGATTTTAAAGCGGGGTTGTTGGTATTTAGATGAATTAAGAGCGCGTC	14907
QY	323	ThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPheAlaAsp	342
Db	14908	ACAATTTGGCAGCAACAATATCTGTTCAAGAAATCTGATGCCAGCTTAATTTACTCCG	14967
QY	343	ArgAlaIleAlaLeuAspHisGluGluGlyCys-----CysTyrLeuLeuAlaLeu	359
Db	14968	CATTTTTTGTCTTCGATCATCACAGAGTAACTTTATGAAAGTTTATACACCAAAAT	15022
QY	360	AspArgArgGlyHisAspAspGlyAlaArgAlaTyrLeuArgGluThrAlaGluThrLeu	379
Db	15028	GGCGAGCCACATCAT-----TGTCACCGGAACTTTCGTTACCAAT	15066
QY	380	ThrGlyLeuAlaValAlaArgAlaProAlaGluProThrProAlaMetValPheGlyIlePro	399
Db	15070	ACA-----AAAAATACCGGTGAAGTAGTACACTTTATTCCTGGCGCGCTGT	15117
QY	400	GluAlaIleAlaGlyPheGlyProLeuAlaArgAlaArgHisAspAspAlaTyrLeu	419
Db	15118	GATCAAAATAAG-----ATTGTCTGGAATATGGCGCTGGAATATATAG	15167
QY	420	LeuArgIleAspGluCysLeuLeuGluIleArgHisGlyGluSerTyrGluIleCysLeu	439
Db	15163	GATGGAAATCCATAATACATCTCAATACATCACAGATGGTGAAGTCTATGAAATCTGTG	15222
QY	440	ThrAsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuArg	459
Db	15223	ACGAATAGGGCCAAAGATGACTTATTCAGACATCCATTCATGACGCTTATCGCGCTAGTGC	15282
QY	460	AlaIleSerProValProTyrGlyAlaLeuLeuGluPheProGluLeuSerValLeuSer	479
Db	15283	CATGCCAGTCTGTACCTTATGGCGCTTATCTGGCTGTGGGAATTTTCTGTACTCAAT	15342
QY	480	AlaSerProGluArgPheLeuThrIleGlyAlaAspGlyValGluSerTyrProIle	499
Db	15343	GCATCACCAAGAAAGTTTTTACGATGATGAGAGACCCCAAAATTTAGATCCAGACCGATT	15402
QY	500	LeuGlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuAlaAspLeu	519
Db	15403	AAGGGGACCCCGCCACCGGMAAAACCGGACCGCATGATCAGACATTTGGCAGATGAAATGG	15462
QY	520	AlaGlyArgGluValAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAsp	539
Db	15463	AGTCACTTACTTAAGATAGGGCAAAAATCTGATGATGTGACTTAAAGCGCCACTGAT	15522
QY	540	LeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGluThr	559
Db	15523	CTTATATACGATATGTCGTCACAGTAAAGTGCATGTACTGAATTTATTTAAAGTTGAGAC	15582
QY	560	TyrAlaProValHisGluLeuValSerThrIleArgGlyArgLeuArgProGlyThrSer	579

Dd		15853	TTCTCTTGTGATCAATTGGTTTGACCATTCGCGGCATATTATGTGAAGAACGTCA	15642
Oy		580	ThrlAalAcysValArgAlaalapeProGlySerMetThrGlyValProLySlys	599
Dd		15643	TCCATPAGAAGGTATCGGGCATGTTTTCCAGAGAGATCATATGACGGGCCACGAAGAAAG	15700
Oy		600	ArGrHmcGlullleIeaSPargLeuGIuglGlyProArgGlyValYrSerGlyAla	619
Dd		15703	CGCAGATGAGATTAATGTGATGCTTAGAGTCATCGCCTGTGGTGTACTCCGGTCGG	15762
Oy		620	LengIYTPheAlaleuSerGlyAlaalAspleSerlleVallleArgHrrlleVal	639
Dd		15763	CTTGAGTGGTATCTTTCACGTGGGAGAGTCGAATTAGCATTTGTGTAGCACCGCGGA	15822
Oy		640	LeuAlaspgIylGlnalagIupheGlyValGIylGlyAlalleValserLeuSerAepGln	659
Dd		15823	TTACATCAAGAGTCGTGTAATTCGGATTCGGGGGGCAATTCCTCATTCATCTATCCT	15882
Oy		660	GIUGlUGlUpheHrgIuthrValVallyAla	670
Dd		15883	TATGCGCAATTGGAAGACATTCGTCAAAGCG	15915
<b>RESULT 12</b>				
ID	ACF71773	standard; DNA; 2019 BP.		
XX	ACF71773;			
DT	20-NOV-2003	(first entry)		
DE	Photorhabdus luminescens nucleotide sequence #10240.			
KM	Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;			
KW	detection; food; gene expression; plant; animal; microorganism; toxin;			
KM	antibiotic; biopesticide; virulence factor; disease model; plague;			
KW	whooping cough; gene; ds.			
OS	Photorhabdus luminescens.			
PN	WO200294867-A2.			
PD	28-NOV-2002.			
Pf	07-FEB-2002; 2002MO-IJB003040.			
PR	07-FEB-2001; 2001FR-00001659.			
PA	(INSP ) INST PASTEUR.			
PA	(CNRS ) CNRS CENT NAT RECH SCI.			
PI	Duchand E, Taouric S, Glaeser P, Frangeul L, Kunst F, Danchin A,			
Buchrieser C;				
WP1;	2003-148459/14.			
<b>Genomic sequence of Photorhabdus luminescens and encoded polypeptides,</b>				
<b>useful e.g. as therapeutic antimicrobials and agricultural pesticides.</b>				
<b>Claim 2; SEQ ID NO 10240; 1205bp; French.</b>				
<b>The invention relates to the isolation of genes and their encoded</b>				
<b>proteins from Photorhabdus luminescens. The isolated sequences are</b>				
<b>sources of probes and primers for detecting the genome of P. luminescens</b>				
<b>and related species; to study polymorphisms, for gene analysis and for</b>				
<b>detection/amplification of the genes. Antibodies (Ab) raised against the</b>				
<b>polypeptides encoded by the genes are used for detection/identification</b>				
<b>of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that</b>				
<b>carry a gene-containing vector are used to select compounds that</b>				
<b>modulate, regulate, induce or inhibit expression of the genes in plants,</b>				
<b>animals or microorganisms other than P. luminescens and are able to alter</b>				
<b>response or sensitivity to toxins and antibiotics produced by P.</b>				
<b>luminescens. Cells transformed to express the genes are useful for</b>				



recombinant production of the proteins, particularly toxins and antibacterials useful as insecticides, bactericides and fungicides. The genes, proteins, vectors containing the genes and Ab are also useful therapeutically (to treat microbial infection by bacteria or fungi that are sensitive to P. luminescens-encoded toxins or antibiotics) and as biopesticides. Other uses of the genes and the proteins are as virulence factors and for identifying targets of human diseases for which P. luminescens is a model (particularly plague and whooping cough). This sequence represents one of the isolated P. luminescens genes

XX Sequence 2019 BP; 531 A; 386 C; 522 G; 580 T; 0 U; 0 Other;

#### Alignment Scores:

Score:	2,57e-90	Length:	2019
Percent Similarity:	1400.50	Matches:	301
Beet Local Similarity:	57.76%	Conservative:	86
Query Match:	44.93%	Mismatches:	246
	39.20%	Indels:	37
	10	Gaps:	6

US-10-089-514-2 (1-686) x ACF71773 (1-2019)

```
QY 22 GIUAIAThGlyGlnProProValValProAspAlaAspTrpSerArgLeuPro 41
   |||||
Db 7 GAAGTACCGGCGGTGCGCGTCCGTAGTACCACTTTGACTTATGAAGAGTAAAG 66
   |||||

QY 42 ValGluAspPheAspAlaIleValValSerProGlyProGlySerProAspArgValArg 61
   |||||
Db 67 ATAGAGAAATATGATCGGATGCTTTCTCCGGGGCCAGAGACATCGGCTGAACAGT 126
   |||||

QY 62 AspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValIleuGlyValCys 81
   |||||
Db 127 GATTGGGGGTCTGCTGAGTATTTGCACATGCCAAGTCCACTGTTGGGTATTGT 186
   |||||

QY 82 LeuGlyHisGlnGlyIleAlaGlnLeuPheGlyValValGlyLeuAlaProGluPro 101
   |||||
Db 187 CTCGGTATCAGGGGATTAACCTGCTTTGGCGGCTTGAGAACATGCCCAATCCG 246
   |||||

QY 102 MetHisGlyArgValSerGluValArgHisThrGlyGluAspValPheArgGlyLeuPro 121
   |||||
Db 247 GTACATGCTTATCGCATCGCATCAGAACACTGGAGAAAGATTTCTCGGTTGGCT 306
   |||||

QY 122 SerProPheThrAlaValArgTyrHisSerLeuAlaAlaThrAspLeuProAspGluLeu 141
   |||||
Db 307 GAACAGTTTGAAGTGTGATATCACTGTTGATGTAACCGGTTGCCGACATCTG 366
   |||||

QY 142 GluProLeuAlaTrpSerAspArgGlyValValMetGlyLeuArgHisArgGlyLeuPro 161
   |||||
Db 367 AAATGTACTCGCCGAGCTGATGATGCTTAATCATGCAATTGAGCACACCGAAGCGCG 426
   |||||

QY 162 LeuTrpGlyValGlnPheHisProGlySerIleGlySerAspPheGlyArgGluIleMet 181
   |||||
Db 427 ATCTGGGGGTTCAATTTCATCTGATGATGATTTGAAATATGGGCGTGACTATCT 486
   |||||

QY 182 AlaAspPheArgAspLeuAlaLeuAlaHisIleArgAlaArgArgHis----- 197
   |||||
Db 487 AGTAATTTGGTGGCGATAGCGAGAGAGCTGTAATAAGAAATATGAACAGAAATTCACCGT 546
   |||||

QY 198 -----GlyAlaAspSerProTyr 203
   |||||
Db 547 CAGAAATATATGTAATCATATGATGAGCTTATCTTGCATGAGTGAAGACAAAGATCTC 606
   |||||

QY 204 GluLeuHisValArgArgValAspValLeuProAspAlaGluGluValArgArgGlyCys 223
   |||||
Db 607 AACCTGAATATATCGTAATGTAAAGAGATGATCGATCTGAATCCCTGTTTATCAACGT 666
   |||||

QY 224 LeuProGlyGluGlyThrThrPheTrpLeuAspSerSerSerValLeuGluGlyAlaSer 243
   |||||
Db 667 TATGGAAATGATACGATGCTTTCTGGCTTGATAGAAATTCATATGACCAATGCC 726
   |||||

QY 244 ArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTyrIleuThrTyrArgVal 263
   |||||
Db 727 CGCTTCTCATGATGGGCGAGCGGTATGAACAAAGGTGCTGAGCTTGAATATCACTGTA 786
   |||||
```

```
QY 264 AlaAspGlyValValSerValArgGlySerAspGlyThrThrArgThrArgPro 283
   |||||
Db 787 CAGACTGGAATTTGGCTTTGATCGTCTCTGATGGCGAGAGATT---CTTCAGGAGAT 843
   |||||

QY 284 PhePheAspTyrLeuGluGluGluGluGluArgArgValProValaProGluLeu 303
   |||||
Db 844 TTTTTCGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 903
   |||||

QY 304 ProPheGluPheAspLeuGlyTyrValGlyTyrLeuGlyTyrGluGluValaGluThr 323
   |||||
Db 904 CCATTGGTATTTAAAGCGGCTTGTGTGTATTTTAAAGATATGAATTAAGGCGCTGACA 963
   |||||

QY 324 ThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPheAlaAspArg 343
   |||||
Db 964 ATTTGGCAGAACAAATATGCTTCAAGACATCTGATGCCAGCTTAATATTACTCCGCAAT 1023
   |||||

QY 344 AlaIleAlaLeuAspHisGlnGluGlyCys-----CysTyrLeuLeuAlaLeuAsp 360
   |||||
Db 1024 TTTTTCGCTGATCATCAGCAGAGTAAACTTTATGAAATGTTTATCACCAATATAGG 1083
   |||||

QY 361 ArgArgGlyHisAspAspGlyValaArgAlaTrpLeuArgGluThrAlaGluThrThr 380
   |||||
Db 1084 CAGCCACATCAT-----TGTCACCGGAACCTTTCGTTACCAATACA 1125
   |||||

QY 381 GlyLeuAlaValaArgAlaProAlaGluProThrProAlaMetValPheGlyIleProGlu 400
   |||||
Db 1126 -----AAAATACCGGTGAAGTGAATCTGACTTATTTCTGGCGCTGTGAT 1173
   |||||

QY 401 AlaAlaIleGlyPheGlyProLeuAlaArgAlaArgHisAspLeuAspAlaTyrLeuLys 420
   |||||
Db 1174 CAAATATAG-----ATTGTCTGAATATGGCGTGAATATATATGAT 1218
   |||||

QY 421 ArgIleAspGluCysLeuLysGluIleArgAsnGlyGluSerTyrGluIleCysLeuThr 440
   |||||
Db 1219 GGAATCCATTAATCACTCCAAATACATCAAGATGTGTGATCTTAATGAATCTGTGTAG 1278
   |||||

QY 441 AsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuArgAla 460
   |||||
Db 1279 AATTAGGCCCAAGATGACTTATTCAGACATCATTTGAGACGCTTATTCGGGTATGCTCAT 1338
   |||||

QY 461 IleSerProValProTyrGlyValaLeuLeuGluPheProGluLeuSerValLeuSerAla 480
   |||||
Db 1339 GCCAGTCTGTACCTTATAGCGCTTATCTGCGCTGGGGAATTTTCTGTACTGATGCA 1398
   |||||

QY 481 SerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValaGluSerLysProIleLys 500
   |||||
Db 1399 TCACCAAGAAACGTTTTTACGTATTGATGACGCCAGAAATATTGAGTCCAGACCGATTAG 1458
   |||||

QY 501 GlyThrArgProArgGlyGlyTyrThrAlaGluGluAspGlyValArgLeuValaAspLeuAla 520
   |||||
Db 1459 GGGACCCGGCCACGGGAAAAACGGCAGCCATGATCCAGATTTGACATGGAAATTTGACT 1518
   |||||

QY 521 GlyArgGlyLysAspArgAlaGluAsnLeuMetIleValaAspLeuValaArgAsnAspLeu 540
   |||||
Db 1519 CAGTCACTCAAGATAGGCGCAAAAATCGATGATAGTGAATTTAGTGGCCATGATCTT 1578
   |||||

QY 541 AsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValaGluThrTyr 560
   |||||
Db 1579 AATCAGGTATGTCTGCCAGGTAGTGTGATGATGATGATTTTAAAGTTGAGAGCTTC 1638
   |||||

QY 561 AlaProValHisGlnLeuValSerThrIleArgGlyValaArgLeuArgProGlyLysSerHis 580
   |||||
Db 1639 TCTTCTGATCATCAATTTGGTTTCGACCAATACCGGCTCAATTTTGAAGAAACGTCATCC 1698
   |||||

QY 581 AlaAlaCysValaArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLysLysArg 600
   |||||
Db 1699 ATAGAGGCTATCCGGGCAATGTTTCCAGAGAGATCAATGACCGGCGCACAAAGAACGCC 1758
   |||||

QY 601 ThrMetGluIleIleAspArgLeuGluGluGluGluProArgGlyValaTyrSerGlyAlaLeu 620
   |||||
Db 1759 ACGATGAGATTAATGATGCTTATGAGTATGATCGGCTCGTGGTGTACTCCGTCGCGCTT 1818
   |||||
```



OY	652	GIYTPPheAlaLeuSerGlyAlaAlaapLeuSerIleValIleArgTrpIleValLeu	640
		... ..	
Db	1819	GGATGGTATCTTCTTCAGTGGTGAAGTGAAGATGAGCATTTGTGTAGCACC	1877
OY	641	AlaapGlyGlnAlaGluPheGlyValGlyGlyAlaIleValSerLeuSerAspGlnGlu	666
Db	1879	CATCAAGAGTGGTGCTTAATTCGGCATTTGGGGGGGCATTTGTTCTCTATCTGATCCTTAT	1933
OY	661	GIUGluPheThrGluThrValIlyVala	670
Db	1939	GCGGAATTTGAAGAGACATTTGCTCAAAAGC	1968
RESULT 13			
ADB06077			
ID	ADB06077	standard; DNA; 2046 BP.	
XX			
AC	ADB06077;		
XX			
DT	20-NOV-2003	(first entry)	
XX			
DE	Alloicoccus oclitis antigenic protein encoding DNA SEQ ID NO:17.		
XX			
KW	Alloicoccus oclitidis; antigenic protein; immunogenic; immunisation;		
KW	gene therapy; Gram-positive bacterium; infection; gene; de.		
XX			
OS	Alloicoccus oclitis.		
XX			
PN	MO2003048304-A2.		
PD	12-UN-2003.		
XX			
PF	25-NOV-2002; 2002MO-US036123.		
XX			
PR	29-NOV-2001; 2001US-0333777P.		
XX			
PR	18-NOV-2002; 2002US-0426742P.		
XX			
PA	(AMHP ) WYETH HOLDINGS CORP.		
XX			
P1	Fletcher LD, Mcmichael JC, Russell DP, Zagursky RJ;		
DR	WPI, 2003-505284/47.		
XX			
PT	P-PSDB; ADB06078.		
XX			
PT	New Alloicoccus oclitis polynucleotides and polypeptides, useful for		
PT	treating and diagnosing diseases, drug screening assays and monitoring of		
PT	effects during drug clinical trials.		
XX			
PS	Claim 7; SEQ ID NO 17; 1019pp; English.		
XX			
CC	The present invention describes an isolated polynucleotide (1) of		
CC	Alloicoccus oclitis genomic DNA, which encodes an antigenic protein.		
CC	Alloicoccus oclitis is a Gram-positive bacterium. Also described: (1)		
CC	an isolated polypeptide that is encoded by the polynucleotide (1); (2) an		
CC	expression vector comprising the novel isolated polynucleotide (1); its		
CC	complement, degenerate variant or fragment; (3) a genetically engineered		
CC	host cell, transfected, transformed or infected with the vector of (2);		
CC	(4) an antibody specific for the polypeptide of (1); (5) an immunogenic		
CC	composition comprising the polypeptide, its complement, biological		
CC	equivalent or fragment, or the polynucleotide that is comprised in the		
CC	expression vector; (6) a pharmaceutical composition comprising the		
CC	polypeptide of (1) and a carrier; (7) a protein chip comprising an array		
CC	of the polypeptides of (1), their biological equivalent or fragment; (8)		
CC	immunising against Alloicoccus oclitis by administering to a host the		
CC	immunogenic composition; (9) detecting and/or identifying Alloicoccus		
CC	oclitidis in the biological sample; (10) a kit comprising a container		
CC	containing the novel polynucleotide, its degenerate variant or fragment,		
CC	or the antibody of (4); and (11) producing a polypeptide by culturing the		
CC	genetically engineered host cell under conditions suitable to produce the		
CC	polypeptide from the culture. (1) can be used in gene therapy. The		
CC	polynucleotides, polypeptides, antibodies and compositions of the present		
CC	invention can be used for treating and diagnosing diseases, drug		
CC	screening assays and monitoring of effects during drug clinical trials.		
CC	The polynucleotides are useful for expressing and detecting Alloicoccus		

[illegible]











Db	6211	ATGAGATGAGAGAGACCTACTCACCACCTCCACGAGCTGTTACACGGTTTACAGCCGGGTC	6152
Qy	575	ArgProGlyThrSerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThr	594
Db	6151	AAAGAGACCTAGATGTCGTGTTGAAGTGTAAAGAAAAACACCTTCCCGGGCGGGTCCATGACT	6092
Qy	595	GlyAlaPArgProGlySerArgThrMetCnilelleaspArgLeuGluGluGlyProArgGly	614
Db	6091	GGGGGGCTTAAAAAAGAACCCCTGGAAATTTATGTATGACTTAAGAGCAGTTCCAGAGGG	6032
Qy	615	ValTyrSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleVal	634
Db	6031	ATCTATTTCGGACATCATTTGGCTTTTATGACCACCAATTCCTACTATAGATCTTAACATGCT	5972
Qy	635	IleArgThrIleValIleuAlaAspGlyGlnAlaGluPheGlyValGlyAlaIleVal	654
Db	5971	ATCCGGACACGATGTCGTGAGAGACAGACAAGGACAGCATCGGCGGGGGTGCATTGTC	5912
Qy	655	SerLeuSerArgPngIngluGluGluPheThrGluThrValIleValValArgAlaMetVal	674
Db	5911	ATGCTATCTGATCTCGAAGAAAGAGTTGATGAACTTTTAAAGCTAAGGGGCGCTTG	5852
Qy	675	ThrAlaIleu 677	
Db	5851	TCCGCTTG 5843	
RESULT 15			
ID	ADB95077	standard; DNA; 2760 BP.	
AC	ADB95077;		
DT	04-DEC-2003	(first entry)	
DE	A. thaliana gene 55483	#SEQ ID 75.	
XX	Plant; herbicide; weed; crop field; growth; development; gene; ds.		
XX	Arabidopsis thaliana.		
PN	WO2003008440-A2.		
XX	30-JAN-2003.		
PD	16-JUL-2002;	2002WO-EP007929.	
PF	16-JUL-2001;	2001US-0305806P.	
PR	20-FEB-2002;	2002US-0358416P.	
XX	(SYGN ) SYNGENTA PARTICIPATIONS AG.		
PA	Levin JZ, Patton DA, McElver JA, Budziszewski CJ, Zhou Q, Aux GW;		
PI	Tosberg J, Weglich Glover L, Ashby CS, Thomas CR, Madhavan E;		
PI	Lewis S, Dunn J, Cates E, Law MD;		
XX	WPI; 2003-229557/22.		
DR	P-PSDB; ADB95078.		
PT	Identifying an herbicidal compound, useful for controlling undesirable		
PT	vegetation, comprises combining a polypeptide with a compound to be		
PT	tested for the ability to bind to the polypeptide or inhibit the activity		
PT	of the polypeptide.		
XX	Example 9; SEQ ID NO 75; 273pp; English.		
CC	The invention relates to a method for identifying a herbicidal compound.		
CC	The method of the invention comprises combining a polypeptide having at		
CC	least 90% identical to any one of 48 69-1008 residue amino acid sequences		
CC	(designated as PI-P48), given in the specification, with a compound to be		
CC	tested for the ability to bind to the polypeptide or inhibit the activity		
CC	of the polypeptide, under conditions conducive to binding or inhibiting,		
CC	respectively. Also disclosed is a method for killing or inhibiting the		
CC	growth or viability of a plant by applying to the plant the herbicidal		

[illegible]



```

Db 970 TATGATGTCAGTATTAAATTCAAGACATACTGAAATTATGGAGTCGTC 1029
Qy 192 -----HisArgAlaArgArgHisGlyAlaAspSerProTyrGluLeuHisValArgArg 209
Db 1030 AAATTCATACATCCCTCGCTCGAAGAAACATTAATGACACTGCAAAACATGACAGTG----- 1083
Qy 210 ValAspValLeuProAspAla-----GluGluValArgArgGlyCysLeuPro 225
Db 1084 -----CCTGATGCTACTCAATGCTGAAAGAACTTCTGAACTAGATGTACA 1131
Qy 226 GlyGluGlyThr----- 229
Db 1132 GGAATGCTGCTTCTAGCTATTTTGGGAACCTTAAGCTCTGTTTCTGCCAAGACAAATGCT 1191
Qy 229 ----- 229
Db 1192 GTAGAGCTCTTGATATGATGATTCATCATATCAAAACACATACAAATTTGCTGAG 1251
Qy 229 ----- 229
Db 1252 TTGAAATGCAAGACATGACGCTTTGGCATTAAGTTGGTGAAGTAAATATATTT 1311
Qy 230 -----ThePheTyrLeuAspSerSerVal 238
Db 1312 ATGGAATCTTTGGCAAGATAGAGAAATGATACCTTTTGGCTGATCTTCTTCTAGT 1371
Qy 239 LeuGluGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTyr 258
Db 1372 GACAAATGCTAGACGATTTTCTTTCATGCGCGGTAAAGTGATCTCTGGAAGCA 1431
Qy 259 LeuThrTyrArgValAlaAsp-----GlyValValSer 269
Db 1432 TTGACATTTTATGTTATCTGATCAAAAGTAGGTTACATCAAAACATGCGGACATCTTCTG 1491
Qy 270 ValArgGlySerAspGlyThrThrThrArg-----ThrArgProPhePheAsnTyr 287
Db 1492 ATTGAAAGATTCTCAGAGTTCTACTGAGAAACAAATTCCTGGAAGAGCTTTCTTGATTT 1551
Qy 288 LeuGluGluGlnLeuGluArgArgValProValAla-----ProGluLeuProPhe 305
Db 1552 CTCGGAAGAGCTTCATCTATCTATGATGAGAGGACTTCGAAGAGTTGCCCTTT 1611
Qy 306 GluPheAsnLeuGlyTyrValGlyTyrGluGlyTyrGluLeuValAlaGluThrThrGly 325
Db 1612 GATTTTGGCGGTGAATACGTAGGTGTGATGGGTATGATTAAGTGAA---TGTCGA 1668
Qy 326 AspPro-----AlaHisArgSerProHisProAspAlaAlaPheLeuPheAlaAspArg 343
Db 1669 ATGCCAATTAAATGCTCAAAATCCAAACGCTCCAGATGCAATGTTCTTTTGGCGAATAT 1728
Qy 344 AlaIleAlaLeuAspHisGlnGluGlyCysCysTyrLeuLeuAlaLeuAspArgArgGly 363
Db 1729 GTTGCGCATTCATCATCAACTCGATGACGTTTATATATATTCGCTTACGAAGGGA 1788
Qy 364 HisAspAspGlyAlaArgAlaTyrPheAspArgGluThrAlaGlu-----ThrLeuThr 380
Db 1789 -----ACTGCAGAAACCTCTTCCGTAATGATACGAAGAGAGCTCATTTAGCTTGATG 1842
Qy 381 GlyLeuAlaValArgAlaProAlaGluProThrProAlaMetValPheGlyIleProGlu 400
Db 1843 GGTTTGTCCAGAAAGATTGGAGATCAAACT---CTTCCAGTTATAGATTCACTCCAA 1899
Qy 401 AlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHisAspLysAspAlaTyrLeuLys 420
Db 1900 TCCAAACAAAGTTTGTCT-----GACAAATCCGAGAGCAGTATATCAAC 1947
Qy 421 ArgIleAspGlyCysLeuLysGluIleArgAsnGlyGluSerTyrGluIleCysLeuThr 440
Db 1948 GATGTTCAAGCGTGTATGAAGATATCAAGAGCGGAGCTACGAGCTTTGTCTCACC 2007
Qy 441 AsnMetValThrAlaProThr---GluAlaThrAlaLeuProLeuTyrSerAlaLeuArg 459

```

```

Db 2008 ACTCAAAACGAGAAAGAAATAGGAATGCTGATCTTTGGGACTTATCTCCACCTGAGA 2067
Qy 460 AlaIleSerProValProTyrGlyAlaLeuLeuGluPhe-----ProGluLeuSerVal 477
Db 2068 GAGAGGAATCCAGCACATATGACGATTTCTCAACTCTCAAAATCCAAATCTGCTCTTA 2127
Qy 478 LeuSerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGluSerLys 497
Db 2128 TGCTCTTCGTCCTTAAAGAGTTTCTTAAGCTGGACAGAAATGGAATGCTTGAAGCAAG 2187
Qy 498 ProIleLysGlyThrArgProArgGlyGlyTyrThrAlaGluGluAspGluArgLysVal 517
Db 2188 CGGATTTAAGGTACTATAGCTGCTGCTCCACGCTCGAAGAGATGAATTTCTTAATG 2247
Qy 518 AspLeuAlaGlyArgGlyLysAspArgAlaGluAsnLeuMetIleValAspLeuValArg 537
Db 2248 CAATTGAAATCAGTGAAGAAATCAAGCCGAGAAATCTGATGATGTTGACTTCAAG 2307
Qy 538 AsnAspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluVal 557
Db 2308 AATGATCTCGCTGCTGTGTGAGCCTGCTCAGTCCATGTATCACTCATGATGATGA 2367
Qy 558 GluThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGly 577
Db 2368 GATCATACACAACAGTATCATCAATGCTGAGCAGATCCGTGACTGAAAAAACAAGAT 2427
Qy 578 ThrSerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaPro 597
Db 2428 ATTAGTCCAGTGAAGATGTAAAGCTGCTTCCCTGCGGTTCAATGACTGGTGCCCA 2487
Qy 598 LysLysArgThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTyrSer 617
Db 2488 AAACTAAGATCTGTGAGATTTCTCGAATCTCTAGAGAACTGTTGAGAGGCTTTTACTCT 2547
Qy 618 GlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThr 637
Db 2548 GGTCAATCGGATTTCTCGTATATGTTGACTTGTGATCTGAATTTGTGATTAAGAAC 2607
Qy 638 IleValLeuAlaAspGlyGlnAlaGluPheGlyValGlyAlaIleValSerLeuSer 657
Db 2608 GTAATTAATCATGAAGATGAGCTTCATTGGAGCAGAGAGAGCTATTGTTGCATTATCA 2667
Qy 658 AspGlnGluGluPheThrGluThrValValValArgAlaMetValThrAlaLeu 677
Db 2668 AGTCCAGAGATGATTTGAGAAATGATTTTAAGACTAGAGCTCTCTAATGACATC 2727

```

Search completed: October 5, 2005, 10:48:02  
 Job time : 1745.67 secs



**This Page Blank (uspio)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 10:07:08 ; Search time 375.417 Seconds  
(without alignments)  
2989.973 Million cell updates/sec

Title: US-10-089-514-2  
Perfect score: 3573  
Sequence: 1 MRTLIDNDSFTNLFXYI.....VVKARAMVTALDGSAAVAGAR 686

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues  
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/usptool\_h/US10089514/runat\_04102005\_105746\_8071/app\_query.fasta\_1.1621  
-Database=Patents.NA -OPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi  
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10089514@cgn2\_1\_116 @runat\_04102005\_105746\_8071 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents.NA.\*  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/prodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/prodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1816	50.8	2220	3	US-08-765-907A-14
2	1816	50.8	2220	4	US-09-987-614A-14
3	1096.5	30.7	4496	4	US-08-765-907A-6
4	1096.5	30.7	4496	4	US-09-987-614A-6
5	725.5	20.3	2888	4	US-08-765-907A-1
6	725.5	20.3	2888	4	US-09-987-614A-1
7	623.5	17.5	7760	4	US-09-987-614A-1
8	615	17.2	1308	4	US-09-902-540-845
9	615	17.2	1431	4	US-09-902-540-8203
10	578.5	16.2	1431	4	US-09-252-991A-13733
11	570.5	16.0	4403765	3	US-09-252-991A-13413
12	570.5	16.0	4411529	3	US-09-103-840A-2

13	563	15.8	1506	4	US-09-252-991A-811	Sequence 811, App
14	556	15.6	1503	4	US-09-328-352-1525	Sequence 1525, Ap
15	556	15.6	645	2	US-08-403-852D-9	Sequence 9, App11
16	536	15.6	645	3	US-08-510-646B-9	Sequence 9, App11
17	556	15.6	645	3	US-09-231-818-9	Sequence 9, App11
18	556	15.6	645	4	US-09-635-359B-9	Sequence 9, App11
19	550	15.4	1518	4	US-09-489-039A-6879	Sequence 6879, Ap
20	548.5	15.4	816	4	US-09-252-991A-13943	Sequence 13943, A
21	510	14.3	2079	4	US-09-248-796A-1387	Sequence 1387, Ap
22	509	14.2	1428	4	US-09-543-681A-336	Sequence 336, Ap
23	494.5	13.8	1392	4	US-09-328-352-1856	Sequence 1856, Ap
24	494	13.6	54484	4	US-09-902-540-1272	Sequence 1272, Ap
25	487.5	13.6	1542	4	US-09-540-236-53	Sequence 53, App1
26	487.5	13.6	49617	4	US-09-596-002-48	Sequence 28, App1
27	485.5	13.6	1473	4	US-09-902-540-5760	Sequence 5760, Ap
28	474	13.3	1664976	4	US-08-916-421B-1	Sequence 1, App11
29	474	13.3	1664976	4	US-09-692-570-1	Sequence 1, App11
30	468	13.1	10357	3	US-08-961-527-191	Sequence 191, App
31	465	13.0	1464	4	US-09-583-110-373	Sequence 373, App
32	465	13.0	1464	4	US-09-107-433-751	Sequence 751, App
33	457.5	12.8	2684	4	US-09-252-991A-12457	Sequence 12457, A
34	449.5	12.6	1722	4	US-09-583-110-2581	Sequence 2581, Ap
35	449.5	12.6	1725	4	US-09-107-433-1608	Sequence 1608, Ap
36	449.5	12.6	13926	3	US-08-961-527-5	Sequence 5, App11
37	439	12.3	3065	3	US-09-199-637A-129	Sequence 129, App
38	438.5	12.3	1734	3	US-09-486-382B-12	Sequence 12, App1
39	437.5	12.2	1734	3	US-09-486-382B-10	Sequence 10, App1
40	434.5	12.2	1821	3	US-09-486-382B-10	Sequence 10, App1
41	433	12.1	1650	2	US-09-001-826-24	Sequence 24, App1
42	433	12.1	1650	4	US-09-264-854-24	Sequence 24, App1
43	430	12.0	2161	2	US-09-001-826-4	Sequence 4, App1
44	430	12.0	2161	2	US-09-001-826-25	Sequence 25, App1
45	430	12.0	2161	4	US-09-264-854-4	Sequence 4, App11

## ALIGNMENTS

RESULT 1  
US-08-765-907A-14  
Sequence 14, Application US/08765907A  
Patent No. 6352839  
GENERAL INFORMATION:  
APPLICANT: BLANC, Veronique  
APPLICANT: THIBAUT, Denis  
APPLICANT: BAWAS-JACQUES, Nathalie  
APPLICANT: BLANCHE, Francis  
APPLICANT: COUZET, Joel  
APPLICANT: BARRIERE, Jean-Claude  
APPLICANT: DEBUSSCHE, Laurent  
APPLICANT: FAMECHON, Alain  
APPLICANT: PARIS, Jean-Marc  
APPLICANT: DUTROC-ROSSET, Gilles  
TITLE OF INVENTION: Streptogramins And Method For Preparing Same By  
FILE REFERENCE: Streptogramin genes  
CURRENT APPLICATION NUMBER: US/08765, 907A  
CURRENT FILING DATE: 1997-03-20  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 14  
LENGTH: 2220  
TYPE: DNA  
ORGANISM: Streptomyces pristinaespiralis  
US-08-765-907A-14

Alignment Scores:  
Pred. No.: 3,33e-177  
Score: 1816.00  
Percent Similarity: 64.03%  
Best Local Similarity: 53.47%  
Query Match: 50.83%  
DB: 3  
Length: 2220  
Matches: 385  
Conservative: 76  
Mismatch: 207  
Indels: 52  
Gaps: 11



US-10-089-514-2 (1-686) x US-08-765-907A-14 (1-2220)

QY 1 MetArgThrLeuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGlnTyrIle 20  
Db 70 GTGCGAACCTGTGTGATCGCAACTACGACTCTTACCTACCAACTCTTCCAGATGCTG 129  
QY 21 GlyGluAlaThrGlyGlnProProValValProAsnAsp-----AlaAspTyrPser 38  
Db 130 GCCGAGGTGAAACGGCCCGCTCGCTCGTCCGCAACGACGACACCGCACTGGGACG 189  
QY 39 ArgLeuProValGluIleAspPheAspAlaIleValValSerProGlyTyrProGlySerProAsp 58  
Db 190 GCCCTGACGGCGGCGCACTTGCACCAAGTGTGTCTCACCGCGCCCGGCGCACTCCGCG 249  
QY 59 ArgGluArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeu 78  
Db 250 ACCGACACCGAAGCTGGGCTCTGACCAAGTGTGTCTCACCGCGCCCGGCGCACTCCGCTG 309  
QY 79 GlyValCysLeuGlyIleGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAla 98  
Db 310 GGGGTGTGCTGGGCGCACGAGCCCTGTGCTGCTGCGCGCGCGCGCGCTGTCTCCACGCA 369  
QY 99 ProGluProMetHisGlyArgValSerGluValArgHisThrGlyGlyLeuAspValPheArg 118  
Db 370 CCGGAACCTTTTACGCGCGCACACGACGACATCCGCGCACGCGGCGCGCTGTTCGCG 429  
QY 119 GlyLeuProSerProPheThrAlaValArgTyrHisSerLeuAlaAlaThrAspLeuPro 138  
Db 430 AACATCCCTCCCTCCGCGTACCGGTGTCTGTTACCATCTCGGTGACCGTCCGCGCACTGCC 489  
QY 139 AspGluLeuGluProLeuAlaTyrPserAspAspGlyValValMetGlyLeuArgHisArg 158  
Db 490 GCCGACCTGCGCGCACCGCGCACACCGCGGACGCGGACGTGATGCGCGCTGTCCGCGCAC 549  
QY 159 GluIleProLeuTyrPylValGlnPheHisProGluSerIleGlySerAspPheGlyArg 178  
Db 550 CACCTGCGCGCTGTGCGGTGCGAGTTCACCCGGAATCGATCGACGGAACAACGCGCAC 609  
QY 179 GluIleMetAlaAsnPheArgAspLeuAlaLeu-----AlaHisHisArg----- 193  
Db 610 CGGATGCTCGCCAACTTCCGCGACCTGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 669  
QY 194 AlaArgArgHisGlyValAspSerPro----- 202  
Db 670 ACCGACGCAATACCGCACCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 729  
QY 203 -----TyrGluLeuHisValArgArgValAspValLeuProAsp 215  
Db 730 GCGTCCGCGCGGTGGGAGATACCGGCTGCATGTGCGGAGGTGCGCTGCGCGCGCGAC 789  
QY 216 AlaGluGluValArgArgGlyCysLeuProGlyGluGlyThrThrPheTyrPheAspSer 235  
Db 790 GCGGACCGCGCTTACACCGCTGTGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 849  
QY 236 SerSerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeu 255  
Db 850 AGCGCGGTGAGCGCGGCGCTGCGCGCTTCACTTCTGCGCGCGCGCGCGCGCGCGCGCG 909  
QY 256 AlaGluTyrLeuThrTyrArgValAlaAspGlyValValSerValArgGlySerAspGly 275  
Db 910 GCGGAACAATCACTCAACGATCGCGCACCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 963  
QY 276 ThrThrThrArgThrArgArgPro-----PhePheAsnTyrLeuGluGlnLeuGlu 293  
Db 964 TCGAGGCGGAGACCG 1023  
QY 294 ArgArgArgValProValAlaProGluLeuProPheGluPheAsnLeuGlyTyrValGly 313  
Db 1024 GCCCGCGCGCTGCC---GCCACCGGCGCTGCGCTTGAAGTTCAACCTCGGCTACGTCGG 1080  
QY 314 TyrLeuGlyTyrGluLeuLysAlaGluThrThrGlyAspProAlaHisArgSerProHis 333  
Db 1081 TACCTGCGGTACGAGACCAAGCGCGACGCGCGCGGAGGACGCGCACCGCGCGGAACTG 1140

QY 334 ProAspAlaAlaPheLeuPheAlaAspArgAlaIleAlaLeuAspHisGlnGluCys 353  
Db 1141 CCGAGCGGCGCTTCAATGTTCGCCGACCGGAGTGTGCGCTGACCGACGACGAGCGGCGG 1200  
QY 354 CysTyrLeuLeuAlaLeuAsp-----ArgArgGlyHisAspAspGlyValArgAla 370  
Db 1201 GCGTGGCTCTGGCACTGAGCAGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260  
QY 371 TrpLeuArgGluThrAlaGluThrLeuThrGlyLeuAlaValArg----- 385  
Db 1261 TGGCTACCGAGCGCGCGCGGACCTTCGCGCACCGCGCGCGCGCGCGCGCGCGCGCTTAC 1320  
QY 386 AlaProAlaGluProThrProAlaMetValPheGlyIleProGluAlaAlaIleArgIlePhe 405  
Db 1321 CTGCGCGAGACCAACTGCCCGCGCTGCAGC----- 1350  
QY 406 GlyProLeuAlaArgAlaArgHisAspIleAspIleAspAlaTyrLeuLysArgIleAspGlyCys 425  
Db 1351 -----GTCACATAACCGCACAGCTGCGCGCTACCGGGAACCTGTGAGGAAATGC 1401  
QY 426 LeuLysGluIleArgAspGlyGlySerTyrGluIleCysLeuThrAspMetValThrAla 445  
Db 1402 CCGCGCTGATACCGACGCGCGAGACTACGAGGTGTGCTTACGAAACATGCTCCGGGTG 1461  
QY 446 ProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuArgAlaIleSerProValPro 465  
Db 1462 CCGCGCGGATGACCGCGCTCACCGCGCTACCGCGCGCTGCGGACCGTCAAGCGCGCG 1521  
QY 466 TyrGlyAlaLeuLeuGluIleProGluLeuSerValLeuSerAlaSerProGluArgPhe 485  
Db 1522 TACGCGCTTACCTGAGTTCCTCCGCGGCGCACCGTGTCTGACTGCTTCAACCGGTTTC 1581  
QY 486 LeuThrIleGlyAlaAspGlyGlyValGlySerLysProIleLysGlyIleThrArgProArg 505  
Db 1582 CTGCGCATGCGCGGAGCGGTGTGGCGGAGTCAAACTATGAAGGACCGCGCGCGCG 1641  
QY 506 GlyGlyThrAlaGluGluAspGlyLysArgLeuArgAlaAspLeuAlaGlyArgGlyLysAsp 525  
Db 1642 GCGCGCGCGCGCGCGCGAGACCGCGCGCTCAAGCGCTCCCTGCGCGCGCGCGAGAAAGAC 1701  
QY 526 ArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAspLeuAsnSerValCysAla 545  
Db 1702 CCGAGCGAACACTGATGATCTGCACCTGTCGCGCAAGCACTCGCGCGCGCGCTGCGAC 1761  
QY 546 IleGlySerValHisValProArgLeuPheGluValGluThrTyrAlaProValHisGln 565  
Db 1762 ATCGGCTTCGTACAGTACCGGCGCTGTTCGAGGTGAGAACCTTACCGCACCGTCCACG 1821  
QY 566 LeuValSerThrIleArgGlyArgLeuArgProGlyThrSerThrAlaAlaCysValArg 585  
Db 1822 CTGCTGACGACGCTCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1881  
QY 586 AlaAlaPheProGlyIleSerMetThrGlyAlaProLysValArgThrMetGluIleIle 605  
Db 1882 GCGCGCTTCCCGCGGCGGTGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1941  
QY 606 AspArgLeuGluGluGlyProArgGlyValTyrSerGlyAlaLeuGlyTyrPheAlaLeu 625  
Db 1942 GACCGGCTCGAAGAGGCGCGCGCGGTGATCTGCGGCGCGCTGCGCTTCACTTCCGCTC 2001  
QY 626 SerGlyAlaAlaAspLeuSerIleValIleArgThrIleValIleAlaAspGlyGlnAla 645  
Db 2002 AGCGCGCGCGCGACTCGACTCTGATCGCACCATGTGTGCCACCGAGAGCGCGCGCG 2061  
QY 646 GluPheGlyValGlyValAlaIleValSerLeuSerAspGlnGluGluIlePheThrGlu 665  
Db 2062 ACCATCGGCGGTGGCGCGCGCTGCGCGCTGTCCGACCGCGAGACGAGGTCCGCGGA 2121  
QY 666 ThrValValLysValArgAlaMetValThrAlaLeuAspGlySerAlaValAlaGlyAla 685  
Db 2122 ATGCTCTTCAAGCGCAGACCACTCGCGCGCTG---CGCGAGGACACGCGGCGCGCG 2178



```

RESULT 2
US-09-987-614A-14
; Sequence 14, Application US/09987614A
; Patent No. 6833382
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIHAUT, Denis
; APPLICANT: BAMS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSET, Gilles
; TITLE OF INVENTION: Stereogramming And Method For Preparing Same By
; TITLE OF INVENTION: Muta-synthesis
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/09/987, 614A
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US/08/765,907
; PRIOR FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 2220
; TYPE: DNA
; ORGANISM: Streptomycetes pristinaespiralis
US-09-987-614A-14

Alignment Scores:
Pred. No.:      3,33e-177      Length:      2220
Score:          1816.00        Matches:     385
Percent Similarity: 64.03%    Conservative: 76
Best Local Similarity: 53.47%   Mismatches: 207
Query Match:     50.83%       Indels:      52
DB:              Gaps:         11

US-10-089-514-2 (1-686) x US-09-987-614A-14 (1-2220)

Oy      1 MetATGThTleuLeuIleAspaSerTYrApsSerPhetThiSaenLeuphegIntYile 20
Db      70 GTGCAGAACCTGTGATTCACAACATAAGCATCCTTAACCAACTCTTCAGATGCTG 129
Oy      21 GIyluaIatThGIyGlnProProValValAlProAsnsp-----AlaSepTrpSer 38
Db      130 GCCGAGGTGAAGAAGCGCCGGCTCCGCTGCTGCTGCCAGACACACACCAGCCGACTGGCAG 189
Oy      39 ArgLeuProValIGluAspPheAspAlaIleValAlSerTrpGlyProGlySerProAsp 58
Db      190 GCCCTTGCGCCGCGGCGCACTTCGACAAAGTCGTGCTCAACCCGCGCCGCGCACCCGCC 249
Oy      59 ArgGIUAvgAAspPheGLyleSerArgAgArGaIAlIeThraSperGLyleuProValLeu 78
Db      250 ACCGACACCGACCTGGGCTCGACGCGCCGCGGGTGATACCGGAATGGAGCCTGCGGTGCTC 309
Oy      79 GIyVALCyLLeuGIyHISglngIyLIleAlaGlneUpheGLyGLyThrValIGyLeuAla 98
Db      310 GGGGTGTGCTGGGCCACACAGGCCCTGTGCTGCTGCCGGCGCCGCGCTGTCCACGCA 359
Oy      99 ProGIUProMetHisGLyArgValSerGIUValArgHisIethrIGlyLusParValPheaArg 118
Db      370 CCGGAACCTTTCAcGGCGCGCACACGACATCCGCCACGAGCGGCGAGGCGCTGTTCGCG 429
Oy      119 GIyLeuProSerProPheThrAlaValArgTyHISerLeuAlalathraSpleuPro 138
Db      430 AACATCTCCCTCCCGGTGACCGTGTGCGGTACCACTCGTAGCCGTCCGGAACCTGCC 489
Oy      139 AserGIULeuGIUProLeuAlatrPserAAspGIyValAlMetGIyLeuAIGHIsarg 158
Db      490 GCCGACCTGGCGGCCACCGCCACACCGCCGACGGGCGAGCTGATGGCGCGTCCGCCACCG 549
Oy      159 GIULvSProlEuTrpGIyValGINphenISproGIUsertIleGIySerAspPheGIyArg 178

```

Db	550	CACCTGCCCCCGCTTGAGGCTGDAAGTTCCACCCCGAATGATTCAGAGGAAACGGAC	609
Qy	179	GIuIlEmEtAlAaSmPhEaRgAsPLeuAlaLeu-----AlaHisHlaRg-----	193
Db	610	CGAAGCTCGGCAACTTCGGGAGCTGTCTCCTGGCGGGGGCGGCGCAACGCCCCCGCAC	669
Qy	194	AlaArgaRgHlsgLyAlaAspSerPro-----	202
Db	670	ACCGAAGCGATACCGCGCACCGCACCCCGCCCCCGCCCCCGCACCGGACCGCC	729
Qy	203	-----TyrGluLeuHisValAlaArgValAspValLeuProHsb	215
Db	730	GCATCCGGCGCGGTGGGGAGATACCGGCTGATGTCGAGAGTCGCTGCGTCGCCAGC	789
Qy	216	AlaGluGluValAlaRgArgGlyCyLeuLeuProGlyGluGlyThrThrPheTyrLeuAspSer	235
Db	790	GCGAAGCGGGCGGTTACACCGCCCTGTGGCCGACAGCGCCCGCGGGTTCTGGCTCGAACAC	849
Qy	236	SerSerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeu	255
Db	850	AGCGCGCTCGAGCGGGCGCTCGCCGCTTCACTTCTCTGGAGCGCCCGCGCGCGCTC	909
Qy	256	AlaGluTyrLeuThrTyrTyrArgValAlaAspGlyValAlaSerValArgGlySerAspGly	275
Db	910	GCGCAACGATACACTTACGAGCTGCGCCGACCGGGCGCGCTGCAAG-----GACGGT	963
Qy	276	ThrThrThrArgThrArgArgPro-----PhePheAnTyrLeuGluGluLeuGlu	293
Db	964	TCAGCGGGCGAGACCCCGCGCGCGCGGACCCCTTTCGACCACTTGAACACGAACTGGCC	1022
Qy	294	ArgArgArgValProValAlaProGluLeuLeuProPheGluPheAsnLeuGlyTyrValGly	313
Db	1024	GCGCGCGCGCTGAGCC---GCCACGGCGCTGCTTTCGATTCAACTCTGGGCTACGTGGC	1080
Qy	314	TyrLeuGlyTyrGlyLeuLeuValGluThrThrGlyAspProAlaHisArgSerProHis	333
Db	1081	TACCTCGGCTTCGAGACCAACGAGCGGACGCGCGCGCGGAGGACGCCACCGCGCGCAACTG	1144
Qy	334	ProAspAlaAlaPheLeuPheAlaAspArgAlaIleAlaLeuAspHisGluGluGlyCys	353
Db	1141	CCGAGCGCGCTTCATATGTTGGCGAGCCGAGTCTCGCTTCGACACAGAAACGAGGGCGG	1200
Qy	354	CysTyrLeuLeuAlaLeuAsp-----ArgArgGlyHisAspAspGlyAlaArgAla	370
Db	1201	GCTGGGCTCTCGGACCTAGAGACGACCGGAGCGCGCGCACCGGACCGCGCGGAAAGC	1260
Qy	371	TyrLeuArgGluThrAlaGluThrLeuThrGlyLeuAlaValArg-----	385
Db	1261	TGGGCTACCGAGCGCGCGCGGACCTCGGCACACCGCGCGCGCGCGCTTCACCTCTG	1320
Qy	386	AlaProAlaGluProThrProAlaMetValPheGlyIleProGluAlaAlaGlyPhe	405
Db	1321	CTGCCCGGACGCAACTGCCCCGCTCGAC-----	1350
Qy	406	GlyProLeuAlaArgAlaArgHisAspIysAspAlaTyrLeuLeuAspGlyCys	425
Db	1351	-----GTCGCACTACCGGCACAGCTCGCGCCGCTACCGGAACTGTGAGGAATGC	1400
Qy	426	LeuLysGluLysLeuAsnGlyGlyLysTyrGluIleCysLeuThrAsnMetValThrAla	445
Db	1402	CGCGCGCTGATCACCGGACGCGGAGAACCTACGAGGTGTGCTCGAGCAACATCTCCGGGCT	1460
Qy	446	ProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuArgAlaIleSerProValPro	465
Db	1462	CCCGGCGGATCGACCGGCTACCGGCTACCGGCGCTCGCGCACCGTAAAGCCCCCGCC	1521
Qy	466	TyrGlyAlaLeuLeuGluPheProGluLeuSerValLeuSerAlaSerProGluArgPhe	485
Db	1522	TACGCGCGCTACCTGCAAGTTCCCGGGGGCACCGTGTCTCAAGCTCTACCCGAAACGGTTC	1581
Qy	486	LeuThrIleGlyAlaAspGlyGlyValGlyLysSerProIleLysGlyThrArgProArg	505



```
Db      1582 CTGCGCATCGCGCGGACGGTTGGCGGAGTCCAAACCCATCAAGGCGACCGCCCGC 1641
Qy      506 G1YGIYThrAlaGluGluAspGluArgLeuArgAlaAspLeuAlaGluArgGluLysAsp 525
      1642 GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1701
Qy      526 ArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAspLeuAsnSerValCysAla 545
      1702 CGCAGCGAGAACCTGATGATCGTCGACCTGTCGCGCAACGACCTCGCGCGCGCGCGCG 1761
Qy      546 IleG1SerValHisValProArgLeuPheGluValGluThrThrAlaProValHisGln 555
      1762 ATGGCTCGGTCCAGACGACCGCGCGCTGTTCGAGGTGAGACCTACCGCACCGTCCACG 1821
Qy      566 LeuValSerThrIleArgGlyArgLeuArgProGlyThrSerThrAlaAlaCysValArg 585
      1822 CTGTCAGACGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1881
Qy      586 AlaAlaPheProGlyLysSerMetThrGlyAlaProLysLysArgThrMetGluIleIle 605
      1882 GCCGCTTCCCGCGCGCGGTGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1941
Qy      606 AspArgLeuGluGluGluProArgGlyValTyrSerGlyValLeuGluTyrPheAlaLeu 625
      1942 GACCGGCTCGAGAGGCGCGCGCGGTGTACTCGGCGCGCTGGGCTGCTTCGCGCGTTC 2001
Qy      626 SerG1ValAlaAspLeuSerIleValIleArgThrIleValLeuAlaAspGlyGlnAla 645
      2002 AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2061
Qy      646 GluPheGlyValGlyAlaIleValSerLeuSerAspGlnGluGluPheThrGlu 665
      2062 ACCATCGCGGTGGCGCGCGCGCGGTGCGCGCTGTCGCGCGCGCGCGCGCGCGCG 2121
Qy      666 ThrValValValAlaArgAlaMetValThrAlaLeuAspGlySerAlaValAlaGlyAla 685
      2122 ATGCTCTCAAGGCGAGACCACTTCGCGCGCTG---CGCGAGGACACGCGGCGCG 2178

RESULT 3
US-08-765-907A-6
; Sequence 6, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSER, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/08/765,907A
; CURRENT FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 4496
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-6

Alignment Scores:
Pred. No.: 2,33e-102 Length: 4496
Score: 1096.50 Matches: 236
Percent Similarity: 57.91% Conservative: 57
Best Local Similarity: 46.64% Mismatches: 162
Query Match: 30.69% Indels: 51
DB: 3 Gaps: 10
```

```
US-10-089-514-2 (1-686) x US-08-765-907A-6 (1-4496)
Qy      1 MetArgThrLeuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGlnTyrIle 20
      3027 GTGCGAACCTCTGTATGACAACTACGACTGGTTCACCTTACCACTTCCAGATGCTG 3086
Qy      21 G1YValAlaThrG1YcInProProValAlaValProAsnAsp-----AlaAspTrpSer 38
      3087 GCGAGGTGAAGGGCGCGCTCGCTCGTCCGGAAGAGACACCGCGACCTGGCAG 3146
Qy      39 ArgLeuProValGluAspPheAspAlaIleValValSerProGlyProGlySerProAsp 58
      3147 GCCCTGGCGCGCGCGCACTTCGAACACGTCGTCTCACCGCGCGCGCGCGCGCGCG 3206
Qy      59 ArgGluArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeu 78
      3207 ACCGACACCGACCTGGGCTCAGCGCGGGGTGATACCGAATGGGACCTGCGCTGCTC 3266
Qy      79 G1YValCysLeuG1YHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAla 98
      3267 GGGGTGTGCTGTGGCCACGAGCGCTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCG 3326
Qy      99 ProGluProMetHisG1YArgValSerGluValArgHisThrGlyGluAspValPheArg 118
      3327 CCCGAACCTTTCACGCGCGCGACGACATCCCGCGCGCGCGCGCGCGCGCGCGCG 3386
Qy      119 GlyLeuProSerProPheThrAlaValArgTyrHisSerLeuAlaAlaThrAspLeuPro 138
      3387 AACATCCCTCCCGCGTGCAGCTGTGCTGCATCCACTCGCTACCGTCCGCGACCTGCC 3446
Qy      139 AspGluLeuGluProLeuAlaTrpSerAspAspGlyValValMetGlyLeuArgHisArg 158
      3447 GCCGACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3506
Qy      159 GlyLysProLeuTrpGlyValGlnPheHisProGluSerIleGlySerAspPheGlyArg 178
      3507 CACCTGCGCGCTGCGCGCGTGCATGCCACCCGAATCATGACGAGACGACGAGCGCAC 3566
Qy      179 GluIleMetAlaAspPheArgAspLeuAlaLeu-----AlaHisHisArg----- 193
      3567 CGGATGCTGCGCAACTTCGCGCGCTGTCTCTGCGCGCGCGCGCGCGCGCGCGCG 3626
Qy      194 AlaArgArgHisG1YAlaAspSerPro----- 202
      3627 ACCGAACGATACCGCGACCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3686
Qy      203 -----TyrGluLeuHisValArgArgValAspValLeuProAsp 215
      3687 GGTCCGCGCGCGGTGGGAGTACCGGCTGATGTGCGGAGGTGCGCTGCGTGC CGAC 3746
Qy      216 AlaGluGluValArgArgGlyCysLeuProGlyGluGlyThrThrPheThrLeuAspSer 235
      3747 GCGAGCGCGCGGTTCACCGCTGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3806
Qy      236 SerSerValLeuGluGlyValAspArgPheSerPheLeuGlyAspAspArgGlyProLeu 255
      3807 AGCGCGGTGACCGCGCGCGCTCGCGCTTACCTTCTGCGCGCGCGCGCGCGCGCGCTC 3866
Qy      256 AlaGluTyrLeuThrTyrArgValAlaAspGlyValValSerValArgGlySerAspGly 275
      3867 GCGGAACAGATCACCTACGACGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3920
Qy      276 ThrThrThrArgThrArgArgPro-----PhePheAsnTyrLeuGluGluGluGlu 293
      3921 TCAGCGCGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3980
Qy      294 ArgArgArgValProValAlaProGluLeuProPheGluPheAsnLeuGlyTyrValGly 313
      3981 GCCCGCGCGCTGCC---GCCACCGCGCTGCGCTTGAATTAACCTGCGCTACGCTGCG 4037
Qy      314 TyrLeuGluTyrGluLeuValAlaGluThrThrGlyAspProAlaHisArgSerProHis 333
```



Dh 4038 TACCTCGGCTACGAGACCAAGCCGACGCGCGGACGAGCGCCACCGCGGAACTG 4097  
Qy 334 ProAspAlaAlaPheLeuPheAlaAspArgAlaIleAlaLeuAspHisGlnGluGlyCys 353  
Dh 4098 CCGGACGGCGGCTTCATGTTGCGGACCGGATGCTCGCCCTTCGACACGAAACAGGGGCG 4157  
Qy 354 CysArgLeuLeuAlaLeuAsp-----ArgArgGlyHisAspAspGlyAlaArgAla 370  
Dh 4158 GCCCGGCTCTGGGACCTGAGCAGCACCAGCGCGCGCCGACCGGACCGCGCGGAAAGC 4217  
Qy 371 TrpLeuArgGluThrAlaGluThrLeuThrGlyLeuAlaValArg----- 385  
Dh 4218 TGGCTACCGACGCGCGCGGACCGCTCCGACCAACCGCGCGCGCGCTTCACCTG 4277  
Qy 386 AlaProAlaGluProThrProAlaMetValPheGlyIleProGluAlaAlaAlaGlyPhe 405  
Dh 4278 CTGCGGACGACCAACTGCGCGCTCGAC----- 4307  
Qy 406 GlyProLeuAlaArgAlaArgHisAspLysAspAlaTyrLeuLysArgIleAspGluCys 425  
Dh 4308 -----GTCACTACCGGACAGCGCTCGCGCTACCGGAACTGTGAGAAATGC 4358  
Qy 426 LeuLysGluIleArgGlnGlySerTyrGluIleCysLeuThrAsnMetValThrAla 445  
Dh 4359 CGCGGCTGTATCACCGACGCGGAGACCTACGAGGTGTGCTGACGAAATGCTCTCGGAGT 4418  
Qy 446 ProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuArgAlaIleSerProValPro 465  
Dh 4419 CCGGCGCGGATCGACCGGCTACCGGCTACCGCGCGCTGCGGACCGTACGCGCGCC 4478  
Qy 466 TyrGlyAlaLeuLeuGlu 471  
Dh 4479 TAGCGCGCTACCTGACG 4496

RESULT 4  
US-09-987-614A-6  
Sequence 6, Application US/09987614A  
Patent No. 683382  
GENERAL INFORMATION:  
APPLICANT: BLANC, Veronique  
APPLICANT: THIBAUT, Denis  
APPLICANT: BAWAS-JACQUES, Nathalie  
APPLICANT: BLANCHE, Francis  
APPLICANT: COUZET, Joel  
APPLICANT: BARRIERE, Laurent  
APPLICANT: DEBUSCHE, Jean-Claude  
APPLICANT: FAMECHON, Alain  
APPLICANT: PAIS, Jean-Marc  
APPLICANT: DUTRUC-ROSSET, Gilles  
TITLE OF INVENTION: Streptogramins And Method For Preparing Same By  
FILE REFERENCE: Streptogramin genes  
CURRENT APPLICATION NUMBER: US/09/987,614A  
PRIOR FILING DATE: 2001-11-15  
PRIOR APPLICATION NUMBER: US/08/765,907  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 6  
LENGTH: 4496  
TYPE: DNA  
ORGANISM: Streptomyces pristinaespiralis  
US-09-987-614A-6

Alignment Scores:  
Pred. No.: 2,33e-102 Length: 4496  
Score: 1096.50 Matches: 236  
Percent Similarity: 57.91% Conservative: 57  
Best Local Similarity: 46.64% Mismatches: 162  
Query Match: 30.69% Indels: 51  
Gaps: 10  
US-10-089-514-2 (1-686) x US-09-987-614A-6 (1-4496)

Qy 1 MetArgThrLeuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGlnTyrIle 20  
Dh 3027 GTGGAACTCTGTCATGACCAACTGACATCTGTCACCTACCAACTCTTTCAGATCTG 3086  
Qy 21 GlyGluAlaThrGlyGlnProProValValProAsnAsp-----AlaAspTyrSer 38  
Dh 3087 GCCGAGGTGAACGGCGCGCTCGCTGCTGTCGCGCAACAGCAGCACACCCGACCTGGCAG 3146  
Qy 39 ArgLeuProValGluAspPheAspAlaIleValValSerProGlyIleProAsp 58  
Dh 3147 GCCCTGGCGCGGCGGCTTCGACCAAGCTGTGCTTCACCGCGCGCGCGCGCGCGCGCG 3206  
Qy 59 ArgGluArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeu 78  
Dh 3207 ACCGACACGACCTCGGCGCTTCACCGCGCGGCTGATCCGAAATGGACCTCGCTGCTC 3266  
Qy 79 GlyValCysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyIleThrValGlyLeuAla 98  
Dh 3267 GGGGTGTGCTGGGCGCACAGCGCGCTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 3326  
Qy 99 ProGluProMetHisGlyValArgValSerGluValArgHisThrGlyGluAspValPheArg 118  
Dh 3327 CCGGAACTTTTACCGCGCGCGCACAGCAGCAGCATCGCGCACAGCGCGCGCGCGCGCG 3386  
Qy 119 GlyLeuProSerProPheThrAlaValArgTyrHisSerLeuAlaIleThrAspLeuPro 138  
Dh 3387 AACATCTCTCTCGCGCTGACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3446  
Qy 139 AspGluLeuGluProLeuAlaTyrSerAspAspGlyValValMetGlyLeuArgHisArg 158  
Dh 3447 GCCGACCTGGCGCGCGCGCGCGCACACCGCGCGGAGCGGAGCTGATGCGCGCGCGCG 3506  
Qy 159 GluLysProLeuTyrGlyValGlnPheHisProGluSerIleGlySerAspPheGlyArg 178  
Dh 3507 CACCTGCGCGCGCTGCG 3566  
Qy 179 GluIleMetAlaAsnPheArgAspLeuAlaLeu-----AlaHisHisArg----- 193  
Dh 3567 CGAGTGTGCGCAACTTCG 3626  
Qy 194 AlaArgArgHisGlyAlaAspSerPro----- 202  
Dh 3627 ACCGAAACGATACCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3686  
Qy 203 -----TyrGluLeuHisValArgArgValAspValLeuProAsp 215  
Dh 3687 GCCTCGCGCGCGGTGGGAGATACCGGCTGCGATGCGCGAGGTGCGCTGCGCGCGAC 3746  
Qy 216 AlaGluGluValArgArgGlyCysLeuProGlyGlyGlyThrThrPheTyrPheAspSer 235  
Dh 3747 GCGGACGCGCGGTTCACCGCGCTGTGCGCGGACCGCGCGCGCGCGCGCGCGCGCGCG 3806  
Qy 236 SerSerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeu 255  
Dh 3807 AGCGGCTGACG 3866  
Qy 256 AlaGluTyrLeuThrTyrArgValAlaAspGlyValValSerValArgIleSerAspGly 275  
Dh 3867 GCGGAACAGATCACTACGACGTCGCGGACCGGCGCGCGCGCGCGCGCGCGCGCGCG 3920  
Qy 276 ThrThrThrArgThrArgArgPro-----PhePheAsnTyrLeuGluGluLeuGlu 293  
Dh 3921 TCAGGCGCGGAGACCG 3980  
Qy 294 ArgArgArgValProValAlaProGluLeuProPheGlnPheAsnLeuGlyTyrValGly 313  
Dh 3981 GCCCGGCGCGCTGCG 4037  
Qy 314 TyrLeuGlyTyrGluLeuLysAlaGluThrThrGlyAspProAlaHisArgSerProHis 333  
Dh 4038 TACCTCGGCTACGAGACCAAGCGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4097



OY		334	ProAspAlaIalpheLeuPheAlaAspArgAlaIalValAlaLeuAspPHISGIngluGLys	353
Dd		4098	CCCGAGGGCGCCTTCATGTGCACGACCGAATGCTCGCCCTTGACACGAACAAGGGCGG	4157
OY		354	CysTyrLeuLeuAlaLeuAsp-----ArgArgGlyHisAspAspGlyAlaArgAla	370
Dd		4158	GCGTCGCTCCGAGCACTGAGCAGCACCCGACGGCCCCGCCACCGCACCGCGCGGAAGCG	4217
OY		371	TTrpLeuArgGluThrAlaGluThrIleuThrGlyLeuAlaValArg-----	385
Dd		4218	TGGCTCACCGACGCGCCGCGGACCCTCCGCCAACACCGGCCCGCCGCGCTTCACCTG	4277
OY		386	AlaProIlaGluProThrProAlaMetValPheGlyIleProGluAlaIalaglyPhe	405
Dd		4278	CTGCCGACGACCAACTGCGCCGCTTGAC-----	4307
OY		406	GlyProLeuAlaArgAlaArgHisAspLysAspAlaTyrLeuLysArgIlaAspGluCys	425
Dd		4308	-----GTTCACTACCGGCACACAGCTGCCTGCCGCTCACGGGAACGTGGAAGAATGC	4358
OY		426	LeuLysGluIleArgAsnGlyGlnSerTyrGluIleCysLeuThrAspMetValThzAla	445
Dd		4359	CGCGCCCTGATACCGACGCGGACGACCTTAACAGAGGTCTCTGACCAATCTCCGGGTG	4418
OY		446	ProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuArgAlaIalSerProValPro	465
Dd		4419	CCCGCGCGGATGACCCGCTCACCGCCTTACCGCGCCCTTGCGACCGTCAGCCCGCCCC	4478
OY		466	TyrglyAlaLeuLeuGlu	471
Dd		4479	TACGCGCGCTCATCTGAC	4496

## RESULT 5

```

US-08-765-907A-1
: Sequence 1, Application US/08765907A
: Patent No. 6352839
:
: GENERAL INFORMATION:
: APPLICANT: BLANC, Veronique
: APPLICANT: THIBAUT, Denis
: APPLICANT: BAMA-S-JACOUES, Nathalie
: APPLICANT: BLANCHE, Francis
: APPLICANT: COUZET, Joel
: APPLICANT: BARRIERE, Jean-Claude
: APPLICANT: DEBUSCHER, Laurent
: APPLICANT: FAMECHON, Alain
: APPLICANT: PARIS, Jean-Marc
: APPLICANT: DUTRUC-ROSSERT, Gilles
: TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
: FILE REFERENCE: Streptogramin genes
: CURRENT APPLICATION NUMBER: US/08/765,907A
: CURRENT FILING DATE: 1997-03-20
: SOFTWARE: PatentIn Ver. 2.0
:
: SEO ID NO 1
:
: LENGTH: 2888
:
: TYPE: DNA
: ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-1

```

Alignment Scores:

Pred. No.:	2,47e-64	Length:	2888
Score:	725.50	Matches:	150
Percent Similarity:	78.70%	Conservative:	20
Best Local Similarity:	69.44%	Mismatches:	45
Query Match:	20.31%	Indels:	1
DB:	3	Gaps:	1

US-10-089-514-2 (1-686) x US-08-765-907A-1 (1-2888)

QY	470	LeuGlupheProGluIuSeSerValIeSerProGluArypHeIuThrIleGly	489
Db	1	CTGCAGTCCCGCGGGGCCACCGTGTAGCTCCACCCGGAACGGTCTCGCATCGAC	60

QY 430 MlaaapglYgIvAlGluSerIySerProIleIySgIyThrArgProaRgIyGIyThrAla 509

Db 61 GGAGCGGCTGGGCGGAGTCCAAACCACATCAAGGGACCGGCCCGCGCGCGGCC 120

QY 510 GluGIuAaSpGIuArgLeuAaRgAlaAaSpLeuAlaGIyArgGIuLyAaPpAaRgAlaIuAa 529

Db 121 GCCCAGAGAGCGCGCGCTCAAGGCTTCCTCGCGCGCGCGAGAAAGAACCGCAGAGAAC 180

QY 530 LeuMetIleValaIaSpLeuValaArgAaAaSpLeuAaSerValaCysAlaIleGIySerVal 549

Db 181 CTGATGATCTGTGACTGTGGTCCGCAACGACCTCGGCGCAGTCTGCGACATCGGCTCGTC 240

QY 550 HisValaProaRgLeuPheGluValaGIuThrYrAlaProValaHisGIuLeuValaSerThr 559

Db 241 CACGTACCGGGCTTTGTGAGGTGAGACATACGCAACGTCACACAGTCTGTGCACACAG 300

QY 570 IleArgGIyArgLeuAaArgProGIyThrSerThrAlaAlaCysValaArgAlaAaPhePro 589

Db 301 GTCCGCGGCGCGCTGGCGGCGCGACGCTCTCCGCGCGCGGTACGGGCGGCTTCCCC 360

QY 590 GIyGIySerMetThrGIyAlaProLyAaRgIyAaRgIyMetGluIleIleAaPArgLeuGIu 609

Db 361 GCGCGGTGATGATACCGGCGCGGCCCAAGGTCGACACATCATGATTCAGCAGCGGCTCGAG 420

QY 610 GIuGIyProaRgGIyValaIyTyrSerGIyAlaLeuGIyTyrPheAlaLeuSerGIyAlaAla 629

Db 421 AAGGCGCGCGCGGGGTGTACTCGGGCGGGGTGGGTCTTGGCTTCAGCGCGCGGCC 480

QY 630 AaPLeuSerIleValaIleArgThrIleValaLeuAaAaSpGIyGlnaIaGIuPheGIyVal 649

Db 481 GACCTCAGCATGTATCTCGCACCATCTGTCCGACCGAGAGAGCGGCCACATCGGCGTG 540

QY 650 GIyGIyAlaIleValaSerLeuSerAaSpGIuGIuGIuPheThrGIuThrValaValy 659

Db 541 GCGCGCGCGGTGTGGCGCTGTCCGACCCGACGACGAGAGTCTCGCAAAATGTCTCTCAAG 6000

QY 670 AlaAaRgAlaMetValaThrAlaLeuAaSpGIySerAlaValaAlaGIyAla 685

Db 601 GGGCAGCACCATCTCGCGGCTCTG---CGCAGGACACAGCGGGCGCC 645

## RESULT 6

```

US-09-987-614A-1
/ Sequence 1, Application US/09987614A
/ Patent No. 683382
/ GENERAL INFORMATION:
/ APPLICANT: BLANC, Veronique
/ APPLICANT: THIBAUT, Denis
/ APPLICANT: BAMA-JACQUES, Nathalie
/ APPLICANT: BLANCHE, Francis
/ APPLICANT: COUZET, Joel
/ APPLICANT: BARRIERE, Jean-Claude
/ APPLICANT: DEBUSSCHE, Laurent
/ APPLICANT: FAMECHON, Alain
/ APPLICANT: PARIS, Jean-Marc
/ APPLICANT: DUTROC-SOSETT, Gilles
/ TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
/ TITLE OF INVENTION: Mltasmasthesis
/ FILE REFERENCE: Streptogramin genes
/ CURRENT APPLICATION NUMBER: US/09/987,614A
/ CURRENT FILING DATE: 2001-11-15
/ PRIOR APPLICATION NUMBER: US/08/765,907
/ PRIOR FILING DATE: 1997-03-20
/ NUMBER OF SEQ ID NOS: 17
/ SOFTWARE: PatentIn Ver. 2.0

```

LENGTH: 2888

```

; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-1

```

Alignment Scores: 2.47e-64 Length: 2888  
Pred. No.:



Score:	725.50	Matches:	150
Percent Similarity:	78.70%	Conservative:	20
Best Local Similarity:	69.44%	Mismatches:	45
Query Match:	20.31%	Indels:	1
Gaps:	4	Gaps:	1

US-10-089-514-2 (1-686) x US-09-987-614A-1 (1-2888)

Qy	470	LeuGIuPherProGIuLeuSerValLeuSerAlaSerProGIuAryPheLeuThrIle61Y	489
Db	1	CTGAGTTCCCGGGGGCAACCGTGTCACTCTCAACCGGAACGGTTCTCGCATCGGC	60
Qy	490	AlaAepGIyGIyValGIuSerIySProlIyGlyYThrAryProAryGIyIYThrAla	509
Db	61	GCGGACGGCTGGGGCGGAGTGTCAAAACCATCAAGGGGACCCGGCCCCGGCGCGCGCCCC	120
Qy	510	GIuGIuAepGIuAryrGIuAryrGAlaAepLeuAlaGIyAryGIuIyAaAaAryAlaGIuAa	528
Db	121	GCCCAAGACCCCGCCGCGAAGGCTCTCCCTGGCGGGCGGAGAAAGACCGGACGAGAAAC	180
Qy	530	LeuWeIleValAaPheValAryAaAaAepLeuAaAaSerValIyGlySerVal	549
Db	181	CTGATGATCGTCAACCTGGTGGCAACGACCTCGGCGAGGTGGAGATGCGTCCGTC	240
Qy	550	HisValProAryrGIuPheGluValGIuThrIYrAlaProValHisGluLeuValSerThr	569
Db	241	CACGTACCGGGCTGTTCGAGGGAGACCTACGGCACCGTTCACACACTGTGACGACG	300
Qy	570	IleAryGIyYrGIuAryrProGIyThrSerThrAlaIaIaCyValAryAlaAlaPhePro	589
Db	301	GTCGCGGGCGGCTGGCGGGCGGACGCTCTCCGCCCCCGGGTACGGGGCGGCTTCCCC	360
Qy	590	GIyGIySerThrIYrAlaProIySlyAryrThrMetGluIleIleAaParyLeuGIu	609
Db	361	GCGGGGTGCATGACCGGGCGGCCCAAGGTCCGACACATGACATGATGACCGGCTCGAG	420
Qy	610	GluGIyProAryrGIyAlYrYzSerGIyAlaIaGIuIYrPheAlaIaLeuSerGIyAlaAla	629
Db	421	AAAGGCCCCGGCGGCTGTACTCTGGGGCGGGCTACTCTTCCCTCAAGGGGGCGGACC	480
Qy	630	AaPLeuSerIleValIleAryrThrIleValIleAaAaAepGIyGluAlaGIuPheGIyVal	649
Db	481	GACCTCAGCATCTGTATCCGACACATCGTGGCGACCGAGAGAGCGGCCACCATCGGCGTG	540
Qy	650	GIyGIyAlaIleValaSerLeuSerAaPGIuGIuGIuIuPheThrGIuThrValValIyS	669
Db	541	GGGGGGGCGCTCGTCCGCTGTCCGACCCGACGACGAGGTCCGGGAAATCTCTCAAG	600
Qy	670	AlaAryAlaMetValThrAlaAaAepGIySerAlaIaValaGIyAla	689
Db	601	GCGGACGACCACTCTCGCGGCTGG---GCGCAGGACACGGCGGGGCGC	645

RESULT 7

; Sequence 845, Application US/09902540

GENERAL INFORMATION:

APPLICANT: Goldman

APPLICANT: Slater

; TITLE OF INVENTION: MYXOCOCCUS

```

/      / 1155 REFERENCE: 38-10 (15849) D
; CURRENT APPLICATION NUMBER: USE

```

PRIOR APPLICATION NUMBER: 60/33

PRIOR FILING DATE: 2000-07-10

; SEQ ID NO 845

1 GENNAME: 7700  
; TYPE: DNA

ORGANISM: *Myxococcus xanthus*  
 HIS-09-903-540-845

**Alignment Scores:**

Pred. No.:	4.25e-53	Length:	7760
Score:	623.50	Matches:	176

Percent Similarity:	50.31%	Conservative:	65
Best Local Similarity:	36.74%	Mismatches:	191

Best Local Similarity:	36.74%	Mismatches:	191
Query Match:	17.45%	Indels:	47
NR:	4	Calls:	14

DB:	4	Caps:	14
-----	---	-------	----

US-10-089-514-2 (1-686) x US-09-902-540-845 (1-7760)

Qy 207 ValArgArgValAspValLeuProAspAlaGluGluValArgArg  
:::|||||

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

[illegible]



```
Dh      3952 ATCGCGGGAAGCGC---CGACGGGGGACCGCCGGAAGAGAGCGCGCTTCTGCACGAG 3896
Qy      519  LeuAlaGlyArgGluYrAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsn 538
Dh      3895 CTGCGACCAACAGAGAGAGCGCGGACGACCGCATCTGTGGACCTTGAGACCGGAC 3836
Qy      539  AspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGlu 558
Dh      3835 GACCTGGCGCGGGGTGACCTTATGCTCGGTGAGTACGAAAGCTGATGACATCTGTC 3776
Qy      559  ThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyThr 578
Dh      3775 GAGTACGCCCAACGCTTCCACATTTGAATCGAAGGTCTCGGCCAGCTCGCGCGGGGTG 3716
Qy      579  SerThrAlaIaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLys 598
Dh      3715 GACCCGCTGACGCTGGTGGGCGCGCTCTTCCAGCGCGGACGATACGGGCGTGGCGGAG 3656
Qy      599  LysArgThrMetGluIleIleAspArgLeuGluGlyProArgGlyValTyrSerGly 618
Dh      3655 ATACGACGATGCAATCATCATCCAGCTGAGCCACATCGCGCGGCTCTATACGGGC 3596
Qy      619  AlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIle 638
Dh      3595 TCGCTGGGTACCTGCTCTTACGCGGTGATTTGATTGAACATCGTCATCCGACGCTG 3536
Qy      639  ValLeuAlaAspGlyGlnAlaGluPheGlyValGlyValIleValSerLeuSerAsp 658
Dh      3535 GTGTGAAAGACCGCGCGGCTTACGCGGAGTGGAGGCGGCATCGTCCACGACTCGGAG 3476
Qy      659  GlnGluGlnGluPheThrGluThrValIleValLysAlaArgAlaMetValThrAlaLeu 677
Dh      3475 CCCCAGAGAAATACAAGAGAGACGCTCAACAGGCGGCTGCGAGCTCTGCGCCCTG 3419

RESULT 8
US-09-902-540-8203
/ Sequence 8203, Application US/09902540
/ Patent No. 6833447
/ GENERAL INFORMATION:
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Wiegand, Roger C.
/ TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
/ FILE REFERENCE: 38-10(15849)B
/ CURRENT APPLICATION NUMBER: US/09/902,540
/ PRIOR FILING DATE: 2001-07-10
/ PRIOR APPLICATION NUMBER: 60/217,883
/ NUMBER OF SEQ ID NOS: 16825
/ SEQ ID NO 8203
/ LENGTH: 1308
/ TYPE: DNA
/ ORGANISM: Myxococcus xanthus
US-09-902-540-8203

Alignment Scores:
Pred. No.:      1,8e-53      length:      1308
Score:          615.00      Matches:      169
Percent Similarity: 50.77%      Conservative: 62
Best Local Similarity: 37.14%      Mismatches:  186
Query Match:    17.21%      Indels:      38
DB:              4          Gaps:          12

US-10-089-514-2 (1-686) x US-09-902-540-8203 (1-1308)
Qy      231  PheTyrPheAspSerSerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAsp 250
Dh      10  TTCCTGCTGAGAGACCGGGTC---AGCGCGAGGGGCGCTATTGTTCTCTCGGC--- 63
Qy      251  AspArgGlyProLeuAlaGluTyrLeuThrTyrArgValAlaAspGlyValSerVal 270
Dh      64  -----GCGAAGCCCTTCTGCTTTCACCGCGAAGAGGAGACAGTCTTCATT 111
```

```
Qy      271  ArgGlySer-----AspGlyThrThrThrArgThrArgArgProPheAsnTyrLeu 288
Dh      112  GATGGCGCCTCCAGACGAGGTACCGCGGAGAGGTG-----CTG 150
Qy      289  GlnGluGlnLeuGlnArgArgValProValAlaProGluLeuProPheGluPheAsn 308
Dh      151  CGCTCGCTCTGGCGCGGTGGCGTGGCGTGCATCCGGAATGCGGCTC---TTCCTCG 207
Qy      309  LeuGlyTyrValGlyTyrLeuGlyTyrGluLeuLysAlaGluThrGlyAspProAla 328
Dh      208  GCGGCGCGGCTGGGCTTCTTCGCTTACGAGCGCGCACCTGTTTGCAGTCCCTGCCCCG 267
Qy      329  HisArgSerProHis-----ProAspAlaAlaPheLeuPheAlaAspArgAlaIle 345
Dh      268  CACCCACGGGAGACGACCTGACACCTGCGCCGAATCGACCTGTCAATTCAACACCTTCG 327
Qy      346  AlaLeuAspHisGlnGluGlyCysTyrLeuLeuAlaLeuAspArgArgGlyHisAsp 365
Dh      328  ACGGTGACCAACACAGAGGCGCAATTGCTCTGTGGCCACGCGGCTCGACTGGAGAGAC 387
Qy      366  AspGlyValArgAlaThrPheLeuArgGluThr---AlaGluThrLeuThrGlyLeuAlaVal 384
Dh      388  TGTGACGCGAGGTGGACCGCTTGGAGGTCCACGTCCGACCGCACCGCCGCGCA 447
Qy      385  ArgAlaProAlaGluProThrProAlaMetValPheGlyIleProGluAlaAlaGly 404
Dh      448  CGGTGCGCA-----CCGCGGATG----- 465
Qy      405  PheGlyProLeuAla-----ArgAlaArgHisAspLysAspAlaTyrLeuLysArgIle 422
Dh      466  GATGGGCGCCACCTCGCGGTATCGCTTCCACCTTCAGTCCGATTCGACGACGCAAG 525
Qy      423  AspGluCysLeuLysGluIleArgAsnGlyLysTyrGluIleCysLeuThrAsnMet 442
Dh      526  GAGCGCGTGGCGGAGTACATCCGCGCGGTGACACGTATCAAGTCAATCTTCGACGCG 585
Qy      443  ValThrAlaProThrGlnAlaThrAlaLeuProLeuTyrSerAlaLeuArgAlaIleSer 462
Dh      586  CTGAGAGTGAACCTTCCGCGTACGCGCTCGGCTCTTACGAGCGGTGCCGACATCAAC 645
Qy      463  ProValProTyrGlyValAlaLeuGluPheProGluLeuSerValLeuSerAlaSerPro 482
Dh      646  CCGGTCCATCTCGCCATCTCGAAGGTGACGCGCTTCACGTCCTCAGCGCTCACCG 705
Qy      483  GlnArgPheLeuThrIleGlyAlaAspGlyValGluSerLysProIleLysGlyThr 502
Dh      706  GAGCGGCTGGTGGTGTG---GAGAACGGCCGAGCCACACGCGGCGCCATCGCGGAGCG 762
Qy      503  ArgProArgGlyGlyThrAlaGluGluAspGluArgLeuArgAlaAspLeuAlaGlyArg 522
Dh      763  CGC---CGAOGGGGACGCGCCGAGAGGAAAGGCGGCTTGTCCACGAGCTGGCGACCAAC 819
Qy      523  GlnLysAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAspLeuAsnSer 542
Dh      820  GAGAAAGAGCGCGCGCAGACGCGCATGCTGTGACCTGAGCGGCAACCACTGGGCGCG 879
Qy      543  ValCysAlaIleGlySerValHisValProArgLeuPheGluValGluThrTyrAlaPro 562
Dh      880  GTGTGACCTTATGCTGTGCTCGAGTGAAGAGCTGATGACATCTCGAGTACGCCAC 939
Qy      563  ValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyThrSerThrAlaAla 582
Dh      940  GTCTCCACATGAATCGAAGTCTGCGCGCACCTGCGCGCGGCTGAGCGCGCTGGAC 999
Qy      583  CysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLysLysArgThrMet 602
Dh      1000  GTGTGCGCGCGCTTTCACAGCGGAGCATTTACGGGCTGCGGAATACGACAGATG 1059
Qy      603  GlnIleIleAspArgLeuGluGluGlyProArgGlyValTyrSerGlyAlaLeuGlyTyrP 622
Dh      1060  CAATCATCACCGAGCTGAGCACATGCGCGCGGCTCTATACGGGCTGCTGGGGTAC 1119
```







```

; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13413
; LENGTH: 1101
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13413

Alignment Scores:
Pred. No.:      8,04e-50      Length:      1101
Score:          578.50      Matches:      137
Percent Similarity: 52.48%      Conservative: 43
Best Local Similarity: 39.94%      Mismatches: 134
Query Match:    16.19%      Indels:      29
DB:              4          Gaps:      4

US-10-089-514-2 (1-686) x US-09-252-991A-13413 (1-1101)

QY      344 AAlaIleAlaLeuSphHisGlnGluGlyCysCysTyrLeuLeu-----AlaLeuAsp 350
      1005 GCGCTGATCGACGACCAACCGCGCGGACCACTGGTATTCCACCGCGATTGGAC 946
QY      361 ArgArgGlyHisAspAspGlyAlaArgAlaTyrLeuArgGlnThrAlaGluThrLeuThr 380
      945 GAGCGGCAAGCGGCGCGCTGATCGCCCTTTCAGCAGAGAAACG----- 901
QY      381 GlyLeuAlaValArgAlaProAlaGluProThrProAlaMetValPheGlyIleProGlu 400
      900 -----CAAGC 895
QY      401 AAlaIleAlaGlyPheGlyProLeuAlaArgAlaArgHisAspIleAspAla-----Tyr 418
      894 GCGCGCGGCAACTTCAGCTTCGGGAAATTCGAGCAGCATCAGCGCATCGCATAT 835
QY      419 LeuIleArgIleAspGluCysLeuIleGluIleArgAsnGlyGluSerTyrGluIleCys 438
      834 CGGAGCGCATTCGCGCGCATCCAGACTACATCCAGCAGGACCACTGCTACAGGTGAAC 795
QY      439 LeuThrAspMetValThrAlaProThrGlnAlaThrAlaLeuProLeuTyrSerAlaLeu 458
      774 TATAGCAGCGCTTCAGCGCGCTTCGACGCGGCTCGCGGCTTCGCGCTTCGCGCTG 715
QY      459 ArgAlaIleSerProValProTyrGlyAlaLeuLeuGlnIlePheProGluLeuSerValLeu 478
      714 CGGAGCGCTTCGCGCGCATTCCTCCGCTACCTGCGCATCGCGCGCGCATCTTC 655
QY      479 SerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGluSerValPro 498
      654 AGCTGTGCGCGGAGCGCTTCCTCAAGCTGAC---AAGGCAAGGTTGAACCCGCGCG 598
QY      499 IleIleGlyIleThrArgProArgIleGlyIleThrAlaGlnGluAspGluArgAlaAsp 518
      597 ATCAAGGAGCACCGCGCGCGGAGAACCCCGAGAGAGACATGGCGCTGGGGGGCTGC 538
QY      519 LeuAlaGlyArgGlyIleAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsn 538
      537 CTCTGTGCGCGCGCGGAGAACCCCGGAAACCTGTGATGCTGCGCTTCGCGCGAAC 478
QY      539 AspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGlnValGlu 558
      477 GACATCGGACGCGATTGCCAACCCTGGAGCGGTACCGAGAGCTGTTCGCGCGGAA 418
QY      559 ThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyIleThr 578
      417 AGCTATCCCAACGTGATCACTGTGTAGCAGCGTCAACCGGCAACTGGGCGGGGAG 358
QY      579 SerThrAlaAlaCysValArgAlaAlaPheProGlyIleGlySerMetThrGlyAlaProIys 598
      357 GAGCGCTTCGACTGTGGAAGGAGCTTCGCGCGGCTCGATCAACCGCGCGCGGAG 298

```

```

QY      599 LysArgThrMetGluIleIleAspArgLeuGlnGlyProArgGlyValTyrSerGly 618
      297 ATTCGGGCATGAGATCATTCAGACAGTGAACGAGCCGAGCGGCACTTACTGCGGC 238
QY      619 AlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIle 638
      237 AGCTGTCTACTTCGACGCTGCGCGGAGATGACAGCTTCGATCGCCATCCGACCCCTG 178
QY      639 ValLeuAlaAspGlyGlnAlaGluPheGlyValGlyIleValIleValSerLeuSerAsp 658
      177 CTGCTCAGGACGCGCGGCTCAGTTCGCGCGCGGCGGCGGATCGCCACATCGCAC 118
QY      659 GlnGlnGluGluPheThrGlnThrValValIleAlaArgAlaMetValThrAlaLeuAsp 678
      117 TGGAGAGACAGATGACAGGAAACCTCGACAAAGTCCGGGTCTGTGGAACCTTGAA 58
QY      679 GlySerAla 681
      57 GGGATGGCC 49
DB

RESULT 11
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC bases
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Alignment Scores:
Pred. No.:      3.65e-43      Length:      4403765
Score:          570.50      Matches:      206
Percent Similarity: 44.14%      Conservative: 99
Best Local Similarity: 29.81%      Mismatches: 285
Query Match:    15.97%      Indels:      101
DB:              3          Gaps:      26

US-10-089-514-2 (1-686) x US-09-103-840A-2 (1-4403765)

QY      48  IleValValSerProGlyProGlySerProAspArgGluArgAspPheGlyIleSerArg 67
      1798378 GTAAGTGAAGACCGGACCTCGGC-----GAATTGGCTGCTTTGCG 1798419
QY      68 ArgAlaIle-----ThrAspSerGlyLeuProValIleGlyValCysLeuGlyHis 84
      1798420 AACGGATCGGTGATGATCCGACCGGCTGGCCTTGACCTGGCAATTTCTTGGCCA 1798479
QY      85 GlnGly---IleAlaGlnLeuPheGlyIleThrValGlyLeuAlaProGluProMetHis 103
      1798480 GTCGCGGAAGTGCGAGGCTTCTTTGGT-----GCAAGCCAGGCGTATCGC 1798524
QY      104 GlyArgValSerGluValAlaArgHis-----ThrGlyGluAspValPheArgGlyLeuPro 121
      1798525 CGCGGATGGAAGAACAGGACCAAGGATCGGTCTCGATGACAGACGCTAAGCCTGCGAG 1798584

```



OY	122	serPro-----PheThrAlaValArgTyHisSer---	131
Db	1798585	AGTCCCGGTCTGATCGGGCAGTTCGAACTCGGATCCGCTGACCCGGTTTCAATATGATGCT	1798644
OY	132	LeuAlaAlaThrAspLeuProAspGluLeuGluProLeu-AlaTrpSerAspArgLy--	150
Db	1798645	CAGGTACACACCGATGGCCGACCTCTTGGCGCGCGCTTCCGGCTGGGGGGTGGCCCCAT	1798704
OY	151	-----ValValMetGlyLeuArg-HisArgGlyLysProLeuT	163
Db	1798705	GCGCGCGCTTTCGCGCGCGGTGATGATGCTGCGGCTTACCGCGGACGATGCGGCTATCT	1798764
OY	163	rrgLyValGlnPheHisProGluSerIleGlySerAspPheGlyValArgGluIleMetAla	183
Db	1798765	GGAAAGATGGTTCCGGTGCACCGCCGACCTCGACGCCACCACTCCGGTAG-----G	1798815
OY	183	snPheArgAspLeuAlaLeuAlaHisHisSArgAlaArgArgHisGlyAlaAspSerProt	203
Db	1798816	ATTTCGGCTCTCTGGCG--GCCAGACACCGGGT-----	1798847
OY	203	yrGlyLeuHisValArgArgValAspValLeuProAspAlaGlu-----GluValA	220
Db	1798848	-----GTTCCGGTGACTGCG--AAGGCTTGGCGGACAGAGAGACCGCGTCCGGCT	1798899
OY	220	rgArgGlyCyLeuLeuProGluGlyGluThrTrpPheTrpLeuAspSerSerValLeuG	240
Db	1798900	ACCGCAACCTCGCGCCCAATCGCCCGGGTACGTTCTCTGTAAGTCGGCCAGAACCGCC	1798959
OY	240	IuGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTrpLeuT	260
Db	1798960	GGTCGTGTCGGGATGCTGCTTATCGGTGCGGG--GCCCAACGGCG--TTGA	1799010
OY	260	hrTyArgValAlaAspGlyValValSerValArgGlySerAspGlyThrThrThrArgT	280
Db	1799011	CCGTGCGCT-----GAGGGGCAAGCGGATGAGCTGGGTCGGTCCGACCAAGACCTCCCA	1799064
OY	280	hrArgArgProPhePheAsnTrpLeuGluGluGlnLeuGlu-----A	294
Db	1799065	CTGGCGGAGACCCGCTCGGGCGCGCTGACGATCTTGAGACTCTGCTGCTACCGCGGATC	1799124
OY	294	rgArgArgValProValAlaProGluLeuProPheGluPheAsnLeuGlyTyTrValGlyT	314
Db	1799125	GTCACTCGGACCGGGCTTCTCCGCGCTG-----TCCGGTGGCATGGTGTGAGTT	1799172
OY	314	yrLeuGlyTyTrpLeuLeuValAspGluThrThrGlyAspProAlaHisArgSerProHis-	333
Db	1799173	TCTTGCGCTATGACATGTGTGGAGCG-----CTGGAAACGATGGCCGAAAC	1799217
OY	334	-----ProAspAlaAlaPheLeuPheAlaAspArgAlaIleA	346
Db	1799218	GGGCGCTGATGATGCTCTGCTCGCGGACATGCTGTTCTTGCCACACGATGTGGCG	1799277
OY	346	IaLeuAspHisGlnGlnGlyCySerGlyTrpLeuLeu-----AlaLeuAspArgArgGlyH	364
Db	1799278	CGGTGATGATCAGCAGAGGGACCATCATCATGTGATGCGCCAACGCGCTGAACGTGAACGCA	1799337
OY	364	IsaAspAspGlyAlaArgAlaTrpLeuArgGlyThrThrAlaGluThrLeuThrGlyLeuAlaV	384
Db	1799338	CCGACGAGCGGGTCT--GACTGGGCGTACAGACACGACCGCGTCTCGGCTCGGACGTATGA	1799394
OY	384	AlaArgAlaProAlaGluProThrTrpProAlaMetVal-----PheGlyTLeProGluAlaA	402
Db	1799395	CCGCGACGCTCGGCGCAACCTACCGTTCACACCGTGGCCACCTTACGCGCAACCGAG---	1799450
OY	402	IaAlaGlyPheGlyProLeuAlaArgAlaArgHisSAspArgValAspAlaTyLeuLysArgI	422
Db	1799451	-----CCGGCGCACCGTGGCGAACCCACCGCTGAAAGAAATATGTGCGATGC	1799496
OY	422	IaAspGlyCyLeuLeuGluGluLeuArgAsnGlyGluSerTyTrpGluTrpCyLeuThrArgH	442
Db	1799497	TGCAATACTTGTTGGATCAAGTTTCAGCCGCGGTGAAGCGTTTCAAGTGTGCTCCCTCGAC	1799556
OY	442	etValThrAlaProThrGluAlaThrAlaLeuProLeuTySerAlaLeuArgAlaIleS	462

[illegible]



## Alignment Scores:

Pred. No.: 3,66e-43 Length: 4411529  
 Score: 570.50 Matches: 206  
 Percent Similarity: 44.14% Conservative: 99  
 Best Local Similarity: 29.81% Mismatches: 285  
 Query Match: 15.97% Indels: 101  
 Gaps: 26

US-10-089-514-2 (1-686) x US-09-103-840A-1 (1-4411529)

```

QY      48  ILValValSerProGluProGlySerProAspArgLysArgAspPheGlyIleSerArg 67
       1807500  GATGTCGCAAAACGCGCACTCGGC-----GATTTGGCTGCTTTGCG 1807541
       68  ArgAlaIle-----ThrAspSerGlyLeuProValIleGlyValCysLeuGlyHis 84
Db      1807542  AACGGGATTCGGTGTGATGCCGACCGGCTGGCTTACCTCGCGCAATTTTGGCCAA 1807601
QY      85  GlnGly---IleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaProGluProMetHis 103
       1807602  GTGCGGAGATGGCAGGCTTTTGGT-----GCAGCCAGGCGTCATGCG 1807646
QY      104  GlyArgValSerGluValArgHis-----ThrGlyGluAspValPheArgGlyLeuPro 121
       1807647  CGCCGGATAGAAACAGAACACGCGGTCGTCGATAGACAGACGCTAACGCTGCGAGG 1807706
QY      122  SerPro-----PheThrAlaValArgTyrHisSer--- 131
       1807707  AGTCCCGGCTGATCGGACGCTGAAAGTCGAGTACCGGTACCGGTGTTTCATAGTCGT 1807766
QY      132  LeuAlaIleThrAspLeuProAspGlyLeuGluProLeu-AlaIleSerAspArgGly-- 150
       1807767  CAGGCTCAACCGGATTCGCCGACCTCTTGCGCGCGCTTGCGGCTGGGGGTGCCCAT 1807826
QY      151  -----ValIleMetGlyLeuArg-HisArgGlyLeuProLeuT 163
       1807827  GCGGCGCGTTTGGCGCGCGTGCATCGTGCAGGCTACGCGCGCGGCGGATCGGCTATCT 1807886
QY      163  ArgValAlaGlnPheHisProGluSerIleGlySerAspPheGlyArgGlnIleMetAla 183
       1807887  GGGAAAGATGGTTCCGCTGACGCGCACTCGCACGCCACCACTCGCGTAG-----G 1807937
QY      183  snPheArgAspLeuAlaLeuAlaHisIleAspArgAlaAspArgHisGlyAlaAspSerProT 203
       1807938  ATTTCGCGCTCTCTGGCG---GCCAGCACCGCGTGG----- 1807969
QY      203  TyrGluLeuHisValArgArgValAspValLeuProAspAlaGlu-----GluValAla 220
       1807970  ----GTTCCGCTGACTCGC---AAGGTCCTGGCCGACAGAGAACCGCGCTGCGGCT 1808021
QY      220  ArgArgGlyCysLeuProGlyGluGluIleThrPheThrPheAspSerSerValLeuG 240
       1808022  ACCGCAAGCTCGCGCAATCGCGCGGAGTACCTGCTGGAATCGGACGCGAGAACGCGC 1808081
QY      240  IuGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTyrLeuT 260
       1808082  GGTGCGGTGCGCATTCGTTTACGTTTACGTTGCGGGG---GCCCAACGCGC-----TTGA 1808132
QY      260  hrTyrArgValAlaAspGlyValValSerValArgGlySerAspGlyThrThrThrArgT 280
       1808133  CCGGCGCT-----GAGGGGCAACGCGGTATGCTGGGTGCGGTGCCCAAGACGCTCCA 1808186
QY      280  hrArgArgProPhePheAsnTyrLeuGluGluGlnLeuGlu-----A 294
       1808187  CTGGCGGAGACCGCGCTCGCGGCGTGCAGTGCATTTGGAGCTGCTGCGGATCGCGGATC 1808246
QY      294  rGArgArgValProValAlaProGluLeuProPheGluPheAsnLeuGlyTyrValAlaGlyT 314
       1808247  GTCAAGTCCGACCGCGGCTTCCGCGCTG-----TCCGGTGGCATGTGCTCGGTT 1808294
QY      314  TyrLeuGlyTyrGluLeuValaGluIleThrThrGlyAspProAlaHisArgSerProHis- 333
       1808295  TCTTGGCTATGACATGTCGCAACG-----CTGGAAAGATTGCCGGAAC 1808339

```

```

Db      1808295  TCTTGGCTATGACATGTCGCAACG-----CTGGAAAGATTGCCGGAAC 1808339
QY      334  -----ProAspAlaAlaPheLeuPheAlaAspArgAlaIleA 346
       1808340  GGGCGCTCGATGACCTCTCCGCGGACATGCTGTGTTGCTGGCGAACGATGTGGCGG 1808399
QY      346  ILeuAspHisGlnGluGlyCysTyrLeuLeu-----AlaLeuAspArgGlyHis 364
       1808400  CGGTGCATCACACAGGCGGACCATTCACGTTATGCTGCCAACCCGCGTAACGTGAACGCGCA 1808459
QY      364  IAspAspGlyAlaArgAlaThrPheArgGluThrAlaGluIleThrGlyLeuAlaVal 384
       1808460  CCAGCAGAGGGGCTC---GACTGGCTTACAGACGACGGCTCCCTGCGTCGACGTGATGA 1808516
QY      384  AlaArgAlaProAlaGluProThrProAlaMetVal-----PheGlyIleProGluAlaA 402
       1808517  CGGACGCTCGCGCCACACATACGTCGACCGGTGGCCACCTTACGCGCACCGGAG--- 1808572
QY      402  IAlaGlyPheGlyProLeuAlaArgAlaArgHisIleAspValAspAlaTyrLeuLysArgI 422
       1808573  -----CCGCGCCACCGTGCAGCAACGCAACGCTCGAAGATATGTTGGTCGATCG 1808618
QY      422  IAspGluCysLeuLysGluIleArgAsnGlyGluSerTyrGluIleCysLeuThrAsnM 442
       1808619  TCGAATATCTTGGTGAATTCAGATTGCAGCGGATGAAGCGTTCCAGGTGTGCCCTCGACG 1808678
QY      442  eValIleThrAlaProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuArgAlaIleS 462
       1808679  GCTTCGAGATGACACCGCATGTCGATCCCATGACGTGATCCGAATTCGCGGGTAAACA 1808738
QY      462  eProValProTyrGlyAlaLeuLeuGluPhePro-----GluLeuS 476
       1808739  ACCCAAGTCCCTACATGATCTACTGACAGTCCGAAATGATGATGGTCAGATGGCACTTTT 1808798
QY      476  eValLeuSerAlaSerProGluIleArgPheLeuThrIleGlyAlaAspGlyValaGluS 496
       1808799  CGATTGTGATTCAGTCTCGGAGGCGCTGTAAAGCTC---CACAAAGCTGGGCGACGA 1808855
QY      496  eLysProIleLysGlyThrArgProArgGlyGlyThrAlaGluGluAspGlyValaGlyLeuA 516
       1808856  CGCATCCGATCCGCGGACACCGGTGGCGCGGAAGACACACAGACAGACGTCGCTTCTGG 1808915
QY      516  rGAlaAspLeuAlaGlyArgGluLysAspArgAlaGluAsnLeuMetIleValAspLeuV 536
       1808916  AAAAAGACTGCTGGCGGACGACAAAGAACGTCGCGAGCATGTGATGCTGTGCACTCG 1808975
QY      536  AlaArgAsnAspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheG 556
       1808976  GCCGAACGACCTGGGTGCGGCTTCGACCGCGGCACTGTTGGGTCCGAGATTACAGCC 1809035
QY      556  IuValGluThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgP 576
       1809036  ACATCGAGCGGTACACACCGTATGATCACCTGTGTCACCGGTGACCGGAAAGCTGGCG 1809095
QY      576  roGlyThrSerThrAlaAlaCysValArgAlaAlaPheProGlyLysSerMetThrGlyA 596
       1809096  AAGGGGACACCGCGGTGACGCGGTGACCGCGCTTTCGCGGCGGACAGCTGTGCGGCG 1809155
QY      596  IAspGlySerArgThrMetGluIleIleAspArgLeuGluGluIleProArgGlyValT 616
       1809156  CGCGGAAGGTGCGCGGAGTGGAGTGAAGAGAGTGAAGAACACCGCGCGGCTTT 1809215
QY      616  YSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleA 636
       1809216  ACGCGGCTGTGCGGTATCTTGAATTCGCGGGAACCTGATCTTGCACATGCGCATTC 1809275
QY      636  rGThrIleValLeuAlaAspGlyGlnAlaGluPheGlyValaGlyAlaIleValSerI 656
       1809276  GCACCGCGGTGATGCTTAACGAGCGCTTATGTCCAGGACGAGCGGTGTGTGTGGCGG 1809335
QY      656  eUeSerAspGlnGluGluPheThrGluThrValValIleValAlaArgAlaMetValThrA 676
       1809336  ACTCAAGGATCTCAAGATACACAGAGCGAGGAAACAAAGCTCGGCTGTGTCACAG 1809395

```



Qy 676 IaleuAspGlySerAlaValAlaGlyAla 685  
 Db 1809396 CGATCGCTGCGCCGACGACGCTGCGCT 1809424

## RESULT 13

US-09-252-991A-811  
 ; Sequence 811, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 811  
 ; LENGTH: 1506  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-811

## Alignment Scores:

Pred. No.: 5,34e-48 Length: 1506  
 Score: 563.00 Matches: 167  
 Percent Similarity: 46.56% Conservative: 77  
 Best Local Similarity: 31.87% Mismatches: 214  
 Query Match: 15.76% Indels: 66  
 DB: 4 Gaps: 15

US-10-089-514-2 (1-686) x US-09-252-991A-811 (1-1506)

Qy 172 ILeuIleuSerAspPheGlyArgGluIleMetAlaSerPheArgAspLeuAlaLeuAlaHis 191  
 Db 52 CTGCCCCCGATGGCTACCAACCGCATCCCGCTGCTTCGAGACCTTGCC----- 102  
 Qy 192 HisArgAlaArgAArgHisGlyAlaAspSerProTyrGluLeuHisValArgArgValAsp 211  
 Db 103 -----GACTTGACACGCGCGCTGCTGATCTACCTGAAGCTGGCCGAC 144  
 Qy 212 ValLeuProAspAlaGluGluValAlaArgArgGlyCysLeuProGlyGluGlyThrThrPhe 231  
 Db 145 GCG-----CGAAGCTCTAC 159  
 Qy 232 TrpLeuAspSerSerSerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAspAsp 251  
 Db 160 CTGCTGAGTCTGCGTACGAGGCGGAGAAATGCGGCGCTATTGATCATCTGCGC----- 213  
 Qy 252 ArgGlyProLeuAlaGluTyrLeuThrTyrArgValAlaAspGlyValValSerValArg 271  
 Db 214 -----CTGCCGTGCGACGCGTGTGCGGAGCTACGACACATCAAGTCGCGATC--- 261  
 Qy 272 GlySerAspGlyThrThrThrArgThrArgArgPhePheAsn----- 286  
 Db 262 AGCATCATGGCATGGAACCGAGCGC-----TTCGATTGCGCGGACCGCGCTG 309  
 Qy 287 ---TyrLeuGluGluGluGluLeuGluArgArgArgValProValAlaProGluLeuProPhe 305  
 Db 310 GCTTTCGTGAGGAGTTCAGAGCGCGCTACCAAGGTGCGCACCGCGCGCTTGCCA--- 366  
 Qy 306 GluPheAsnLeuGlyTyrValGlyTyrLeuGlyTyr-----GluLeu 319  
 Db 367 CGTTTCATGGCGCGCTGCTGCTACTTCGTTTACATGCGTGGCTACGTTGGAATAA 426  
 Qy 320 LysAlaGluThrThrGlyAspProAlaHisArgSerPro-----HisProAspAlaAla 337  
 Db 427 CGCTGGCGACCTGTCCGAACCG-----GACCCGCTGGGCAACCGCGATATCTCTG 477

Qy 338 PheLeuPheAlaAspArgAlaIleAlaLeuAspHisGlnGluGlyCysCysTyrLeuLeu 357  
 Db 478 TTGATGTTGTCGATGCGGATGCGGATGCTATTCGACAACTGGCCGGGAAGATTCACGCCATC 537  
 Qy 358 AlaLeu-----AspArgArgGlyHisAspAspGlyAlaArgAlaTrpLeuArg 373  
 Db 538 GTCTCGCGCATGCTTCCTCCGAGAAATGCTTACGAGCGCGC---GAGCAGCTGTGAG 594  
 Qy 374 GluThrAlaGluThrLeuThrGlyLeuAlaValAlaArgAlaProAlaGluProThrProAla 393  
 Db 595 GAGCTGCTGAGCGCTG-----CGCCAGCGCATCACCGCGCGCTGGCG 639  
 Qy 394 MetValPheGlyLeuProGluAlaAlaGlyPheGlyProLeuAlaArgAlaArgHis 413  
 Db 640 CTGACCTC-----GAGCGCGCCAGAGTCTGAGCGCGCTTCTGCTCAGCTTC 690  
 Qy 414 AspLysAspAlaTyrLeuLysArgGlyLeuAspGlyCysLeuLysGluLeuArgAsnGlyGlu 433  
 Db 691 ACCCGGAGGATCATGAAACGGGTAGAGAGATCAAGGATCATCTCGCGCGCGAC 750  
 Qy 434 SerTyrGluLeuLeuThrAsnMetValThrAlaProThrGluAlaThrAlaLeuPro 453  
 Db 751 TGCATGAGGTGTCGCGCTGCGACGCGCATGCTCATGCAATTCAGCGCGCGCCATCGAC 810  
 Qy 454 LeuTyrSerAlaLeuArgAlaIleSerProValProTyrGlyAlaLeuLeuGluPhePro 473  
 Db 811 CTGTACCGCGCGCTGCGCTGTTTCAATTCGACCGCTTACATGTTCTTCAACTTCGCGC 870  
 Qy 474 GluLeuSerValLeuSerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyGly 493  
 Db 871 GACTTCATGTCCTGCGGACGCTGCGGAGTCTGTGACGGTCTC---GAGATGCGCTG 927  
 Qy 494 ValGluSerLysProIleLysGlyThrArgProArgGlyGlyThrAlaGluGluAspGlu 513  
 Db 928 GTGACCGTGGCCGATTCGCGCGCTACCGCGCGCGGATCAACAAGCGCGGACCTG 987  
 Qy 514 ArgLeuArgAlaAspLeuAlaGlyArgGluLysAspArgAlaGluLeuMetIleVal 533  
 Db 988 GCACTGAGCAGGATGCTGCTGCGACCGCAAGAAATGCGCAGACCTGATGTGATC 1047  
 Qy 534 AspLeuValArgAsnAspLeuAsnSerValCysAlaIleGlySerValHisValProArg 553  
 Db 1048 GACTTGGGCGCAACACGCGGCGGCTGTCGGAATGCGCGCGGTGAAGTCAACCGAA 1107  
 Qy 554 LeuPheGluValGluThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArg 573  
 Db 1108 AAATGCGTATCGAAGCTTATCTCAACGTCATGACATGCTGTCGAACGTCACCGGCA 1167  
 Qy 574 LeuArgProGlyThrSerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMet 593  
 Db 1168 TTCGCGAGAGGCTCAGCGCGATGAGACGCGCTGCGGCGCATTCGCGCGGCGACTCTA 1227  
 Qy 594 ThrGlyAlaProLysLysArgThrMetGluIleIleAspArgLeuGluGluProArg 613  
 Db 1228 TCCGCGCGCCCAAGATCCGCGCATGAGATATGACGAGCTGAGCGCGTCAAGCT 1287  
 Qy 614 GlyValTyrSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIle 633  
 Db 1288 GAGTCTACGCGCGCGCGCTGCTACCTGAGTGAAGCGCAATGACACCGCCGATC 1347  
 Qy 634 ValIleArgThrIleValLeuAlaAspGlyGlnAlaGluPheGlyValGlyAlaIle 653  
 Db 1348 GCCATCCGACCCCGGTGATCAAGAACGTTGAATCTCACGTGAGCGCGCGCGCTATC 1407  
 Qy 654 ValSerLeuSerAspGlnGluGluPheThrGluThrValValLysAlaArgAlaMet 673  
 Db 1408 GTTGCAGACTCGGTGCGCGCGCTGAGTGAAGAAACATCAACAAGCGCGGCGGATG 1467  
 Qy 674 ValThrAlaLeu 677  
 Db 1468 TTCGCGCGCTG 1479

RESULT 14



```
US-09-328-352-1525
; Sequence 1525, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328.352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1525
; LENGTH: 1503
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1525

Alignment Scores:
Pred. No.: 1,75e-47 Length: 1503
Score: 558.00 Matches: 162
Percent Similarity: 47.38% Conservative: 91
Best Local Similarity: 30.34% Mismatches: 203
Query Match: 15.62% Indels: 78
DB: 4 Gaps: 15

US-10-089-514-2 (1-686) x US-09-328-352-1525 (1-1503)

QY 181 MetAlaaspheargleuAlaLeuAlaHis-----HisArgAla 194
D 19 TTAGCGCAATTCCAGCACTTAAAGGCTGCTGTACATACTATTCTCTGTTATCGCA 78
QY 195 ArgAArgHisglYAlaSerSerProTYrGluLeuHisValArgArgValaSerValaLeuPro 214
D 79 CGTTTACGGATACAGAAACCCATTATCTGTTTCTCGTTTAAAGAT----- 129
QY 215 AspaAlaGlulValaArgArgGlyCysLeuProGlyGluGlyThrPheTrpLeuAsp 234
D 130 -----CAAAACACAGGCTTACTTAAATTGAG 153
QY 235 SerSerSerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAspArgGlyPro 254
D 154 TCTGTTAAAGGGGTGAAACTGGGCGCGCTACTCTATGTATGGT----- 198
QY 255 LeuAlaGluTYrLeuThrTYrArgValaAlaAspGlyValaSerValaArgGlySerAsp 274
D 199 TTAGGCAATCCACAGTTTCTCTGTAACGCTGCTTTTATCTATACAAACATGGTGAC 258
QY 275 GlyThrThrThrArgThrArg-----ArgProPhePheAsnTYrLeuGluGluLeu 292
D 259 GGATCAGTAAACACGCAAAATTTGCTAGACCA---TTCCAAATATATCCGTGAATCCAA 315
QY 293 GluArgArgArgValaProValAla-----ProGluLeuProPheGluPheAsnLeu 309
D 316 AAACAATTTTAAAGTTTCTACGCGCTAAACTTACACAGACTTACCA--ACCTTTACGGCG 372
QY 310 GlyTYrValGlyTYrLeuGlyTYrGlu-----LeuVala 321
D 373 GCGTGTGGGTATTTGGGCTACGATGCTGTCCGCTATATCGAGCCACGCTTTAAAGAT 432
QY 322 GluThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPheAla 341
D 433 GTACCTACGCGGTATCCGATT-----ACGCTACCAAGATTATGTTGATGCTCTCA 483
QY 342 AspArgAlaAlaLeuAlaSerHisGlnGluGlyCysGlyTYrLeuLeuAlaLeuAspArg 361
D 484 AAGACACTCATTTGTTTGAACAATCTTAAAGATACCGTATTTTAAATTGTG----- 534
QY 362 ArgGlyHisAspArgGlyAlaArgAlaTrpLeuArgGluThrAla----- 376
D 535 -----CATCGGATTAACAGACAGAGATATGCTTATATAAGCGCTCGCAAAATTAATGAT 588
QY 377 -----GluThrLeuThrGlyLeuAlaValaArgAlaProAlaGluPro---ThrProAla 393
D 589 CAAATTGAACAGTTGTCGCACTCCAGTTAGTTGCAAGCCCAACCATACGCT--- 645
QY 394 MetValPheGlyIleProGluAlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHis 413
D 646 -----CCGCATTTTGAATCAATTAAT 666
QY 414 AspLYsAspAlaTYrLeuLYsArgIleAspGluCysLeuLYsGluIleArgHisGlyGlu 433
D 667 GGTAAAGCAAAATTTCTTAAAGACAGTAGAAGAGTTTAAAGATATATCCGTACAGCGCAT 726
QY 434 SerTYrGluIleCysLeuThrAsnMetValThrAlaProTHrGluAlaThrAlaLeuPro 453
D 727 GTGATCGAGTTGTATCTGCGCAGCGTATGTTGCTGATTGTTGAGAGAGACCTTTACG 786
QY 454 LeuTYrSerAlaLeuArgAlaIleSerProValProTYrGlyAlaLeuLeuGluPhePro 473
D 787 GTTTACCGTCATTTAGCTCATTTAAATCATCATCATCTATCTATTTCTGTTTCAAGCGCA 846
QY 474 GluLeuSer-----ValLeuSerAlaSerProGluArgPheLeuThr 487
D 847 ACGATTACTGATTAATAAACCATTTTCATATTTGCGCTCATACCGGAA--ATTTTATCT 903
QY 488 IleGlyAlaAspGlyGlyValaGluSerLYsProIleLYsGlyTYrThrArgProArgGly 507
D 904 CGTTTAAAGAACGGTATTTGCTACAGTTGCACCTTGCGCAGAACTAGACCGCGGTAA 963
QY 508 ThrAlaGluGluAspGluArgLeuArgAlaAspLeuAlaGlyArgGluLYsAspArgAla 527
D 964 ACTTAAAGAAAGATATATACATTTAGAAAAAGATTTCCTGTCGATGAAAGAGATTGCT 1023
QY 528 GluAsnLeuMetIleValaAspLeuValaArgAsnAspLeuAsnSerValCysAlaIleGly 547
D 1024 GAACATTATATGCTGATGATGATCTTGGCGAAACGATGTAGGCGCGGTATCGAAAAATAGCT 1083
QY 548 SerValHisValProArgLeuPheGluValaGluThrTYrAlaProValHisGlnLeuVal 567
D 1084 AAGTTCCAAGTCACGCGATCAAAATGTGATGTAGCGCTATTCACATGTATGATATTGTT 1143
QY 568 SerThrIleArgGlyArgLeuArgProGlyThrSerThrAlaAlaCysValaArgAlaAla 587
D 1144 TCAATGTTCMAAGGTAAAGTGCCTGATGATATCGATGCATCTTGATATTTAAAGCCACC 1203
QY 588 PheProGlyLYsSerMetThrGlyAlaProLYsArgThrMetGluIleIleAspArg 607
D 1204 TTCCGCGGAGGAAGCTTATCAGGTGCCCAAAATTCGTGCAATGGAAATTTATGTGAA 1263
QY 608 LeuGluGluGlyProArgGlyValaTYrSerGlyAlaLeuGlyTYrPheAlaLeuSerGly 627
D 1264 GTAGAACCTGTGAAAAGGGGAGATTTTGGCGGGCGGTGTATTTGGGATGGCATGGT 1323
QY 628 AlaAlaAspLeuSerIleValIleArgThrIleValaLeuAlaAspGlyGlnAlaGluPhe 647
D 1324 GAAATGGATATATGCTCATTTCCATTCCTGATTGTGTATTCGAGCAAAAGGTATGTGA 1383
QY 648 GlyValaGlyAlaIleValaSerLeuSerAspArgGlnGluGluPheThrGlyThrVal 667
D 1384 CAGCGTGTGCAAGGCTAGTTGTGCTCAATTCAGATTCGAGTGGGAATGAAACTCA 1443
QY 668 ValLYsAlaArgAlaMetValThrAlaLeuAspGlySerAla 681
D 1444 ATTAAAGCTCGCGAGTGTCAAAAGCGTTGATTAATATCATCA 1485
```

```
RESULT 15
US-08-403-852D-9
; Sequence 9, Application US/08403852D
; Patent No. 5891695
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Bianche, Francis
; APPLICANT: Crouzet, Joel
; APPLICANT: Jacques, Nathalie
; APPLICANT: Lacroix, Patricia
; APPLICANT: Thibaut, Denis
```



APPLICANT: Zagorec, Monique  
APPLICANT: Debusche, Laurent  
APPLICANT: De Crecy-Lagard, Valerie  
TITLE OF INVENTION: Polypeptides involved in the  
TITLE OF INVENTION: Biosynthesis of streptogramins, Nucleotide sequences  
TITLE OF INVENTION: Coding for these polypeptides and their use  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fiumegan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/403,852D  
FILING DATE: 10-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR 93/00923  
FILING DATE: 25-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/11441  
FILING DATE: 25-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03806.0054-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4000  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 645 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: S.pristinaespiralis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 61..645  
OTHER INFORMATION: /product= "gene papa"  
US-08-403-852D-9

Alignment Scores:  
Pred. No.: 7,15e-48 Length: 645  
Score: 556.00 Matches: 106  
Percent Similarity: 68.06% Conservative: 24  
Best Local Similarity: 55.50% Mismatches: 59  
Query Match: 15.56% Indels: 2  
DB: 2 Gaps: 1

US-10-089-514-2 (1-686) x US-08-403-852D-9 (1-645)

QY 1 MetAqThrLeuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGlnTyrIle 20  
DB 70 GTGGAACCTGCTGATCGACCACTACGACTCGTTCACTACCACTCTTCAGATGCTG 129

QY 21 GtGluAlaThrGlyGlnProProValValProAsnAsp-----AlaAspTyrSer 38  
DB 130 GCCAGGTGAACGGCGCGCTCGCTCGCGCAAGCAGACACCGCCACCTGGCAG 189

QY 39 ArgLeuProValGluAspPheAspAlaIleValValSerProGlyProGlySerProAsp 58  
DB 190 GCCTGGCGCGCGCGCACTTCGACACGCTGCTCTCACCGCGCGCGCGCACCGCGCC 249

QY 59 ArgGluArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeu 78  
DB 250 ACCGACACCGACCTGGGACCTTCACACCGCGGGTGATACCGAATGGACCTGGCGTGC 309

QY 79 GtValCysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAla 98  
DB 310 GGGGTGTGCTGGGCCACCGAGCCCTGTGCTGTGCTGCGCGGCCGCCCTGTCCACGCA 369

QY 99 ProGluProMetHisGlyArgValSerGluValArgHisThrGlyGlyAspValPheArg 118  
DB 370 CCCGAACCTTTCACGCGCGCACACGACATCCGCGCACGACGCGGAGGCGCTGTTCGCG 429

QY 119 GlyLeuProSerProPheThrAlaValArgTyrHisSerLeuAlaAlaThrAspLeuPro 138  
DB 430 AACATCCCTCCCGCTGACCGGTGTCCTGCTACCTGCTGACCGCTCCGCAACTGCC 489

QY 139 AspGluLeuGluProLeuAlaTrpSerAspAspGlyValValMetGlyLeuArgHisArg 158  
DB 490 GCCGACCTGGCGGCCACCGCCACACCGCGGACGCGGACGCTGATGGCCGTCCGCGC 549

QY 159 GluLeuProLeuTrpGlyValGlnPheHisProGluSerIleGlySerAspPheGlyArg 178  
DB 550 CACCTGCGCGCTGCGCGGTGAGTTCACACCGCAATCGATCGACGCAACGCGCAC 609

QY 179 GluIleMetAlaAsnAspArgAspLeuAlaLeu 189  
DB 610 CGATGCTCGCCAACTTCGCGCGACCTGTGCTG 642

Search completed: October 5, 2005, 21:02:45  
Job time: 6032.42 secs



**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 09:57:48 ; Search time 1509.69 Seconds

(without alignments)  
3137.453 Million cell updates/sec

Title: US-10-089-514-2

Perfect score: 3573

Sequence: 1 MRTLIDNYSFTNHLFOYI.....VKARAWVTALDGSAAVAGAR 686

## Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 7442561 seqs, 3452328358 residues

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODE=frame+g2n.model -DEV=rlh  
-Q=/cg2\_1/USPTO.epool\_R/US10089514/runat\_04102005.105745.8055/app.query.fasta\_1.1621  
-DB=Published.Applications.NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bases -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.csl -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10089514@cgn\_1\_1\_920@runat\_04102005.105745.8055  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXT=7

## Database :

Published Applications NA:\*

- 1: /cg2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cg2\_6/prodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cg2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cg2\_6/prodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cg2\_6/prodata/1/pubpna/PCTUS\_NEW\_PUB.seq:\*
- 6: /cg2\_6/prodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cg2\_6/prodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cg2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cg2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cg2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cg2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cg2\_6/prodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cg2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 14: /cg2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 15: /cg2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cg2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 17: /cg2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 18: /cg2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 19: /cg2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 20: /cg2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 21: /cg2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 22: /cg2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 23: /cg2\_6/prodata/1/pubpna/US11\_NEW\_PUB.seq:\*
- 24: /cg2\_6/prodata/1/pubpna/US11\_NEW\_PUB.seq:\*
- 25: /cg2\_6/prodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 26: /cg2\_6/prodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3573	100.0	2061	20	US-10-472-587-1
2	3127.5	87.5	3305	14	US-10-126-927-68
3	3127.5	87.5	12391	14	US-10-126-927-67
4	1766	44.7	138203	21	US-10-819-386A-1
5	1596.5	44.7	2196	15	US-10-156-761-1167
6	1596.5	44.7	9025608	15	US-10-156-761-11
7	1487	41.6	1860	9	US-09-738-626-1111
8	1487	41.6	2005	21	US-10-494-675-25
9	1487	41.6	3309400	9	US-09-738-626-11
10	1350	37.8	2046	22	US-10-501-282-17
11	1350	37.8	1754382	22	US-10-501-282-1651
12	1210.5	33.9	3023	19	US-10-437-963-3537
13	1068.5	29.9	2523	17	US-10-282-122A-17825
14	1001	28.0	1906	17	US-10-282-122A-16977
15	834	23.3	684707	17	US-10-398-221-9
16	834	23.3	3011208	17	US-10-398-221-9
17	697.5	19.5	1066	18	US-10-424-599-33601
18	623	17.4	1362	17	US-10-282-122A-30135
19	585	16.4	9025608	15	US-10-156-761-1
20	579.5	16.2	1377	17	US-10-282-122A-41081
21	568	15.9	1377	17	US-10-282-122A-41824
22	566.5	15.9	1338	17	US-10-282-122A-32104
23	566.5	15.9	1479	15	US-10-156-761-6148
24	563	15.8	1323	17	US-10-282-122A-19610
25	562.5	15.7	661	20	US-10-425-115-81629
26	561.5	15.7	1365	17	US-10-282-122A-38883
27	561	15.7	1440	17	US-10-282-122A-38883
28	557.5	15.6	1365	17	US-10-282-122A-39512
29	556	15.6	645	20	US-10-716-803-9
30	544.5	15.2	1257	17	US-10-282-122A-33856
31	539.5	15.1	1333	17	US-10-282-122A-23903
32	538.5	15.1	5721	17	US-10-398-221-3850
33	534.5	14.9	1337	17	US-10-282-122A-76794
34	531.5	14.9	495269	17	US-10-398-221-8
35	531.5	14.9	3011208	17	US-10-398-221-2058
36	526.5	14.7	1362	17	US-10-282-122A-6689
37	519.5	14.5	2073	22	US-10-450-763-29714
38	509	14.2	1425	17	US-10-282-122A-32780
39	504	14.1	1284	9	US-09-974-300-482
40	500	14.0	5396	17	US-10-398-221-3849
41	499.5	14.0	1898	22	US-10-450-763-1954
42	498.5	14.0	2256646	19	US-10-470-565-1
43	498	13.9	1518	19	US-10-375-039-15
44	497.5	13.9	2731748	19	US-10-297-465A-1
45	489	13.7	1794	18	US-10-275-026A-173

## ALIGNMENTS

RESULT 1  
US-10-472-587-1  
; Sequence 1, Application US/10472587  
; Publication No. US2004024274A1  
GENERAL INFORMATION:  
APPLICANT: YANAI, Koji  
APPLICANT: SUMIDA, Naomi  
APPLICANT: WATANABE, Manabu  
APPLICANT: MORIYA, Tatsuki  
APPLICANT: MURAKAMI, Takeshi  
TITLE OF INVENTION: Transformants Producing Substances  
FILE REFERENCE: 2003-1302A/MMC/00144  
CURRENT APPLICATION NUMBER: US/10/472,587  
CURRENT FILING DATE: 2003-09-22  
PRIOR APPLICATION NUMBER: 82227/2001



```
; PRIOR FILING DATE: 2001-03-22
; NUMBER OF SEQ. ID NOS: 40
; SOFTWARE: Patentin Ver. 2.1
; SEQ. ID NO. 1
; LENGTH: 2061
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2058)
; US-10-472-587-1

Alignment Scores:
Pred. No.: 0 Length: 2061
Score: 3573.00 Matches: 686
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-10-089-514-2 (1-686) x US-10-472-587-1 (1-2061)

QY 1 MetAArgThrLeuLeuIleAspAanTyrAspSerPheThrHisAsnLeuPheGlnTyrIle 20
Db 1 ATGCGCAGCGCTCTGATCGACAACACTACGACTCGTTCCACCACAACCTGTTCCAGTACATC 60

QY 21 G1G1uua1aThrG1yGlnProProValValProAsnAspAlaAspTTPSerATrLeu 40
Db 61 GGGAGAGCCACCGGGCAACCCCGCTGCTGTCGCCAACGACGCCGATGCTGGCGCTG 120

QY 41 ProValG1uuaSPpheaAPAlaIleValIserProG1ySerProAspAArgG1u 60
Db 121 CCCGTCGAGAGACTTTCACGCGATCGTGTCTCCCGGGCCCCCGACCGCCGACCGGAA 180

QY 61 ArgAspPheG1yIleSerATgATgAlaIleThrAspSerG1yLeuProValIleuG1yVal 80
Db 181 CGGAGACTTTCGGAATCAGCCCGCGGCGATCACCAGCAGCGGCGCTGCGCTCTCGGGGT 240

QY 81 CysLeuG1yHisGlnG1yIleAlaGlnLeuPheG1yG1yThrValG1yLeuAlaProG1u 100
Db 81 CysLeuG1yHisGlnG1yIleAlaGlnLeuPheG1yG1yThrValG1yLeuAlaProG1u 100

QY 241 TGCTTCGGCCACACGGGGCATCGCCACGCTTTCGGCGAACCGTGGCCCTCGCCCGGAA 300
Db 241 TGCTTCGGCCACACGGGGCATCGCCACGCTTTCGGCGAACCGTGGCCCTCGCCCGGAA 300

QY 101 ProMetHisG1yATgValIserG1uValATgHisThrG1yG1uAspValPheaArgG1yLeu 120
Db 101 ProMetHisG1yATgValIserG1uValATgHisThrG1yG1uAspValPheaArgG1yLeu 120

QY 301 CCATGACAGCGCGGCTCTCCAGAGTCCGGCACACCGCGCAGACGCTTTCGGGGGCTC 360
Db 301 CCATGACAGCGCGGCTCTCCAGAGTCCGGCACACCGCGCAGACGCTTTCGGGGGCTC 360

QY 121 ProSerProPheThrAlaValATgTyrHisSerIleuAlaIleThrAspLeuProAspG1u 140
Db 361 CCTTCGCGCTTCAACCGCGCTGCGCTACACTCCCTGGCCGACACGCACTCCCGCAGAG 420

QY 141 LeuG1uProIleuAlaTTPSerAspAspG1yValValMetG1yLeuATgHisATrG1uLys 160
Db 141 LeuG1uProIleuAlaTTPSerAspAspG1yValValMetG1yLeuATgHisATrG1uLys 160

QY 421 CTGGAACCCCTCGCTCGAGCAGACGAGGGGTGCTCATGGGCTTCGCGCACCGCGAAG 480
Db 421 CTGGAACCCCTCGCTCGAGCAGACGAGGGGTGCTCATGGGCTTCGCGCACCGCGAAG 480

QY 161 ProLeuTrpG1yValGlnPheHisProG1uSerIleG1ySerAspPheG1yATrG1uIle 180
Db 161 ProLeuTrpG1yValGlnPheHisProG1uSerIleG1ySerAspPheG1yATrG1uIle 180

QY 481 CCGCTGGGGGCGTCCAGTTCACCCCGAGTCCATCCGACGACCTTTCGGCCGGAGATC 540
Db 481 CCGCTGGGGGCGTCCAGTTCACCCCGAGTCCATCCGACGACCTTTCGGCCGGAGATC 540

QY 181 MetAlaAsnPheATgAspLeuAlaLeuAlaHisHisATrG1aATrATgHisG1yAlaAsp 200
Db 541 ATGGCAACTTCGCGCAGCTCGCCCTCGCCACACCGGGCAGCGGCGCCAGGGGGCGAC 600

QY 201 SerProTyrG1uLeuHisValATrATrG1ValAspValIleuProAspAlaG1uG1uValATr 220
Db 601 TCCCGTACCAACTCCACGTGCGCGCTGACGCTGCGACGCTGCGGACGCGCAAGAGTACGC 660

QY 221 ATrG1yCysLeuProG1yG1yG1yThrPheTrpLeuAspSerSerValIleuG1u 240
Db 661 GCGGCTGCTTGGCGCGAGGAGCAGCAGTTCCTGGCTGAGCAGACCTCGTCTTGAA 720

QY 241 G1yAlaSerATrPheSerPheLeuG1yAspAspAArgG1yProLeuAlaG1yTyrLeuThr 260
Db 721 GGGGCTCGGCTTCTCTCTCTCTCGGAGACGACCGGGCGGCTCGCGGATACCTCACC 780
```

```
QY 261 TyrATrValAlaAspG1yValValSerValATrG1ySerAspG1yThrThrATrThr 280
Db 781 TACCGGTGCGAGCGGCTGATCTTCCTCCGGGCTCCGACGACACCGAGAGC 840

QY 281 ArgATrProPhePheAsnTyrLeuG1uG1uIleuG1uATrATrATrATrATrATrATrATr 300
Db 841 CCGGCGCCCTTCTTCAACTACCTGGAGGAGCAGTCCGAACCGCGACGGGTCCCGTCCGC 900

QY 301 ProG1uLeuProPheG1uPheAsnLeuG1yTyrValG1yTyrLeuG1yTyrG1uLeuLys 320
Db 901 CCGAACTGCGCTTCGAGTTCAACTCGGCTACGTCGCTACCTCGGCTACGACTGAAG 960

QY 321 AlaG1uThrThrG1yAspProAlaHisATrATrATrATrATrATrATrATrATrATrATr 340
Db 961 GCGGAGACCAACCGGCGACCCCGGACCGGTCCTCCGACCCCGGACCGGCTTCTTTC 1020

QY 341 AlaAspATrAlaIleAlaLeuAspHisGlnG1uG1yCysATrTyrLeuLeuAlaLeuAsp 360
Db 1021 GCGACCGGCGCATGCGCTTCGACCAACGAGAAAGGCTGCTTACCTGCTGGCCCTCGAC 1080

QY 361 ATgATrG1yHisAspAspG1yValATrATrATrATrATrATrATrATrATrATrATrATr 380
Db 1081 CCGCGGGGCGACAGACGCGCGCGCTGCTGCTGGAGACGCGCCGACACCTTCACAC 1140

QY 381 G1yLeuAlaValATrAlaProAlaG1uProAlaG1uProAlaMetValPheG1yIleProG1u 400
Db 1141 GGCCTGGCGCTCGCGCCCGCGGCGAGCCGACCCCGCATGCTTTCGGAGTCCCGAG 1200

QY 401 AlaAlaAlaG1yPheG1yProLeuAlaATrATrATrATrATrATrATrATrATrATrATr 420
Db 1201 GGGGGGCGGCTTGGCGCCCTTGCGCGCGCGCGCGACGACGACGAGCGCTTACCTCAAG 1260

QY 421 ATrG1yLeuG1yCysLeuLysG1uIleATrATrATrATrATrATrATrATrATrATrATr 440
Db 1261 CGCATCGACGAGATGCTTCAGAGAGATCCGCAACCGGAGTCTTACGAGATTCGCTGAC 1320

QY 441 AsnMetValThrAlaProThrG1uAlaThrAlaLeuProLeuTyrSerAlaLeuATrATr 460
Db 1321 AACATGTACACGCGCGGACCGAGGCGAGCGGCTGCGCTTACTCCGCTCGCGGCC 1380

QY 461 IleSerProValProTyrG1yAlaLeuLeuG1uPheProG1uLeuSerValIleuSerAla 480
Db 1381 ATCAGCCCGCTCCGTAAGCGCGCCCTGCTCGAGTTCGCCGAATCGGTGCTGAGCGCC 1440

QY 481 SerProG1uATrPheLeuThrIleG1yAlaAspG1yG1yValG1uSerLysProIleLys 500
Db 1441 TCGCCCGAGCGGTTCTTCAGATCGCGCGCGACGCGCGCTGAGTCCAAAGCCATCAAG 1500

QY 501 G1yThrATrATrATrATrATrATrATrATrATrATrATrATrATrATrATrATrATrATr 520
Db 1501 GGGACCGGCGCGCGGCGGCGACCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560

QY 521 G1yATrG1uLysAspATrATrATrATrATrATrATrATrATrATrATrATrATrATrATr 540
Db 1561 GCGCGGAGAAAGAACCGGCGCGAGAACTGATGATGTGACTGTGCGCAGACAGCTTC 1620

QY 541 AsnSerValCysAlaIleG1ySerValHisValProATrLeuPheG1uValG1uThrTyr 560
Db 1621 AACAGGCTGCGCGCATCGGCTCCGTCACGTCGCGCGCTCTTCGAGGTGAGACCTTAC 1680

QY 561 AlaProValHisGlnLeuValSerThrIleATrATrATrATrATrATrATrATrATrATr 580
Db 1681 GCGCGCTGACACAGTGTGTGTGACCACTCGGGAGCGGCTCGCGCGCGCGCACAGCAC 1740

QY 581 AlaAlaCysValATrATrATrATrATrATrATrATrATrATrATrATrATrATrATrATr 600
Db 1741 GCGGCTGCTTACGCGCGCGCTTCCCGGCGGCTCCATACCGGCGCGCGCGCAAGAGCC 1800

QY 601 ThrMetG1uIleIleAspATrLeuG1uG1yProATrATrATrATrATrATrATrATrATr 620
Db 1801 ACCATGAGATCATGACCGCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
```



Oy	62	GLTYRphhAlaleuSerGIyAlaAlaApLeuSerIleValIleArGThrIleValIleu	64
Dh	1861	GGATGTTCCGCTTCCACCGCGCGCCGCGACCTTCAGCATGTCAATCGACATGCTG	197
Oy	641	AlaApGIyAlaAlaApLeuPheGIyValGIyGIyAlaIleValSerIleuSerApGIyGlu	660
Dh	1921	GCGCAGCGCCAGGGGAGGTTCCGCGTCCGGGGGCGATGTGTCTCTCCGACCAAGAG	190
Oy	661	GIuGIuPheThrGIuThrValIValIyAlaArgAlaMetValThrAlaIleuApGIySer	688
Dh	1981	GAGAGATTCAACCGAGACCGCTGTAAAGCGCCGCGCATGTATCAGCGCTTCAGCGAGC	20
Oy	681	AlaValAlaGIyAlaArg	686
Dh	2041	GCGTGGCGGCGGCCCA	2058
RESULT 2			
US-10-126-927-68			
: Sequence 68, Application US/10126927			
: Publication NO. US20030082575A1			
: GENERAL INFORMATION:			
: APPLICANT: The Scripps Research Institute			
: APPLICANT: Schultz, Peter G			
: APPLICANT: Wang, Lei			
: APPLICANT: Anderson, John C			
: APPLICANT: Chin, Jason			
: APPLICANT: Liu, David R			
: APPLICANT: Meglery, Thomas			
: APPLICANT: Meggers, Eric L			
: APPLICANT: Mehl, Ryan A			
: APPLICANT: Pastrnak, Miro			
: APPLICANT: Santoro, Stephen W			
: APPLICANT: Zhang, Zhiwen			
: TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids			
: FILE REFERENCE: 54-000120US			
: CURRENT APPLICATION NUMBER: US/10/126,927			
: CURRENT FILING DATE: 2002-04-19			
: PRIOR APPLICATION NUMBER: US 60/285,030			
: PRIOR FILING DATE: 2001-04-19			
: PRIOR APPLICATION NUMBER: US 60/355,514			
: PRIOR FILING DATE: 2002-02-06			
: NUMBER OF SEQ ID NOS: 79			
: SOFTWARE: PatentIn version 3.1			
: SEQ ID NO 68			
: LENGTH: 3305			
: TYPE: DNA			
: ORGANISM: Streptomyces venezuelae			
US-10-126-927-68			
Alignment Scores:			
Pred. No.: 1,3e-301 Length: 3305			
Score: 3127.50 Matches: 618			
Percent Similarity: 91.40% Conservative: 9			
Best local Similarity: 90.09% Mismatches: 42			
Query Match: 87.53% Indels: 17			
DB: 14 Gaps: 3			
US-10-089-514-2 (1-686) x US-10-126-927-68 (1-3305)			
Oy	1	MetArGhrIeuLeuIleAspArTYrAspSerPheThrHisAsnLeuPheGIntYrIle	20
Dh	1	ATGGGCAAGCTTCTGATCGACCACTACGACTGTTTCAACCAAGAACCTGTTCAGTACATC	60
Oy	21	GIyGIuAlaThrGIyGInProProValIValIProAsnAspAlaAsPTIrSerArgIeu	40
Dh	61	GGCAGGCCACCGGCGAGCCGCC---GTCTGTCCCAAGACGCCGACGTGTGGGCTG	11
Oy	41	ProValIuAspPheAspAlaIleValIalSerProGIyProGIySerProAspArgGIu	60
Dh	118	CCCCCTCAGAGACTTCGACCGGATGTCTGTGTCCCGCGGCCCGGACGCCGACCGGAA	17
Oy	61	ArgAspPheGIyIleSerArgArgAlaIleThrAspSerGIyLeuProValIleuGIyAl	80

Db	178	GGGGAATTCCGGAGATTCAGCCGCGGGGGAGATCA	CGAAGCGGCTCGCCGCTCCGAGCTC	237
OY	81	CysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyThrValGlyLeuAlaProGlu		100
Db	238	TGCCTCGGCGCACAGGCGATGCCCCAG-----CTTCGCGCGGAA		276
OY	101	ProMetHisGlyYAspValSerGluValArgHisThrGlyIleAspValPheArgGlyLeu		120
Db	277	CCCAATGACCGCGGGAGTCTCCGAGTCCGGACACCGGCGAGAACGTCCTTCGGGGGCTTC		336
OY	121	ProSerProPheThrAlaValArgTyrHisSerLeuAlaAlaThrAspLeuProAspGlu		140
Db	337	CCCTCGCGGCTTCAACCGCGCGTCCGCTTCAACTCCCTGGCGCGCACGAACTCCCGACGAG		396
OY	141	LeuGlnProIleuAlaATP SerAspAspGlyValValMetGlyLeuArgHisArgGlyLys		160
Db	397	CTCGAACCCCTCGCCTTGAGCGAGACGAGCGGCTGCTCAATGGGCTTCGGCGCACCGGAGAAAG		456
OY	161	ProIleuThrGlyValGlnPheHisProGluSerIleGlySerAspPheGlyValGluIle		180
Db	457	CCGCTGATGGCGGTCGACAGTTCACCGGAGTCCATCGGACGCACTTCGCGCGGAGATGC		516
OY	181	MetAlaAsnPheATPAspLeuAlaLeuAlaHisIleArgAlaArgArgHisGlyValAsp		200
Db	517	ATGGCCAACTTCGCGGAACTCGCCCTGGCGCCACACCGGGGCACTTCGCGACCGGGCGGAC		576
OY	201	SerProTyrGluLeuHisValArgArgValAspValLeuProAspAlaGluGluValArg		220
Db	577	TGGGGCTACAGAACTCCACGTCGCGCGGTGACGTCGTGCGCGGACGCGGAAGAGTATAGC		636
OY	221	ArgGlyCysIleuProGlyGluGlyThrThrPheThrLeuAspSerSerSerValLeuGlu		240
Db	637	CGCCCTCGCTGCGCGCGGAGGGGCGACAGTTCGCGCTGGACACACAGCTTCGCTCCGAA		696
OY	241	GlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTyrLeuThr		260
Db	697	GGCCCTCGCGCTTCTCTCTCTCTGGGAGACACCGCGGCTCGCGGAATCACTCAAC		756
OY	261	TyrArgValAlaAspGlyValValSerValArgGlySerAspGlyThrThrArgThr		280
Db	757	TACGCGCGGCGCGAGCGGTGTCGTCGCGCGGCTCCGACGACACACGACCGGGGAC		816
OY	281	ArgArgProPhePheAsnTyrLeuGlnGluGlnLeuGluValArgValProValAla		300
Db	817	GCGGCGAACCCTTCACCTACCTACGAGAGAGCTGCAACGCGCGCGGATCCGCTGCGCC		876
OY	301	ProGluLeuProPheGluPheAsnLeuGlyTyrValGlyTyrLeuGlyTyrGlyIleuLys		320
Db	877	CCCAACTGCCCTTCGAGATTCACTGAGCTTACCTCGAGTCACTCTCGGCTACGACTGAG		936
OY	321	AlaGluThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe		340
Db	937	GGGAGAACACCGGCGGACCCCGGAGTACCGGCGCGGACCCCGACCGCGGTTCTCTTC		996
OY	341	AlaAspArgAlaIleAlaLeuSerHisGlnGluCysCysTyrGluLeuAlaLeuAsp		360
Db	997	GCCACACGCGCAATCGCCCTGACACACAGAGAAAGCTGCTCACTCGCGGCTCGAC		1056
OY	361	ArgArgGlyHisAspAspArgValAlaArgAlaThrLeuArgGluThrAlaGluThrLeuThr		380
Db	1057	CGCGGGGCGCACGACGAGCGCGCGGCTCGCTGGCGGAGACGGCGGACACTTCAC		1116
OY	381	GlyLeuAlaValArgAlaProAlaGluProThrProAlaMetValPheGlyIleProGlu		400
Db	1117	GGCGTGGCGCGTCCGCGTCCGCGCGAGGCGGACCCCGGCATGCTTCGGGGTCCCGAG		1176
OY	401	AlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHisAspLysAspAlaTyrGluLys		420
Db	1177	GCGCGCGCGGCTTCGCGCCCTCGGCTTCGCGACGCGCACACAGACGAGCGCC-----		1227
OY	421	ArgIleAspGlyCysLeuLysGluIleArgAsnGlyGlySerTyrGluIleCysLeuThr		440
Db	1228	-----TCGGCGCTCCGACGCGGAGATGTGACGATGTGCTTGCCTTAC		1269











QY	100	lu pro met hi s g l y a n g v a l s e g l u v a l a r g h i s t n t g l u s p a l p n e a r g l y	119
Db	18910	GAGCCCGGCTCAGCGGCCACCTCGCGGGTACGGGACACACGGCACCGGCGCTTCCTTCGAGGGG	18968
QY	120	Leu Pro Ser Pro Phe Thr Ala Val Arg Tyr His Ser Leu Ala Ala Thr Asp Leu Pro Asp	139
Db	18970	CTGCCGCGAGCGCGCTGGAGAGTGGTGGTATCCACTTCCTCGCGGTAGCGAACTGCGCCGG	19028
QY	140	Gl u Leu Glu Pro Leu Ala Trp Ser Asp Asp Gly Val Met Gly Leu Arg His Arg Glu	159
Db	19030	GAGCTGGAGGCGCACCGCGCTGGTTCGAGAGACGGGGTCTGTATGGCGCTGCGCCACCGCACGG	19089
QY	160	Lys Pro Leu Thr Pro Gly Val Gln Phe His Pro Glu Ser Ile Gly Ser Asp Phe Gly Arg Glu	179
Db	19090	CTGCCCGCTGGTGGGGGGGTGCAGATTCTCACCCCGAGTCGATCGGCACCGAGACCGCCACCGG	19149
QY	180	Ile Met Ala Ser Phe Arg Asp Leu Ala Leu Ala His His Arg Ala Arg -----	195
Db	19150	CTGCTGGCGCAACTTCGCGACCTACCGGACGCGCACGGCCGACGCGCCGACGCGCCGCGGCGGCGG	19209
QY	196	--Arg His Gly -----	200
Db	19210	GCGGGGACAGGCGACGCTCCGCGCCCGCGCGCGCGCGGAGACGACGGCGCACCGGCG	19269
QY	201	Ser Pro Tyr Glu Leu His Val Arg Arg Val Asp Val Leu Pro -----Asp Ala Glu	217
Db	19270	ACGGCCACGCGCGGCTCCGGGTC--ATCGCCAGTCGGCTGGCCACGCGCTGGAGCGCGGAG	19326
QY	218	Glu Val -----Arg Arg Gly Cys Leu Pro Gly Glu Gly Thr Thr Phe Trp	232
Db	19327	GTCGCGCTTCGACTCGCTGTTCCGACCGGCGGACCACTCC-----TTCTCG	19371
QY	233	Leu Asp Ser Ser Ser Val Leu Glu Gly Val Ser Arg Phe Ser Phe Leu Glu Yasp Asp Arg	252
Db	19372	CTTCGACAGACCGCTCCCGGGGGGAGACTGGGCGAGTCTCCGTATGGGCGACGCGCTCA	19431
QY	253	Gly Pro Leu Ala Glu Tyr Leu Thr Tyr Arg Val Ala Asp Gly Val Val Ser Val Arg Gly	272
Db	19432	GGTCCCTCGCGCCGAGCCGACCGGACCGGACGAGTGCACGCGCGGACCGGTCAACGGTGAGA--	19488
QY	273	Ser Asp Gly Thr Thr Thr Arg Thr Arg Arg Pro Phe Asn Tyr Leu Glu Glu Gln Leu	292
Db	19489	GCCACGCGCGCGGACGACGAGCTGCGAGAGCGGCTTCCTGACCTGGCTGGAGAACGACCTTG	19548
QY	293	Glu Arg Arg Arg Val Pro Val Ala Pro Glu Leu Pro Phe Glu Phe Asn Leu Gly Tyr Val	312
Db	19549	GCGGGGCTGGCGCACCGAGGTG--CCGGAACCTTCCTCGCGGTTCGCGCTCGGCTGGGTC	19605
QY	313	Gly Tyr Leu Gly Tyr Glu Leu Gly Val Glu Thr Thr Gly Yasp Pro Ala His Arg Ser Pro	332
Db	19606	GGCTGGCTGGGCTACGATTGAAGCGGACGACTGGGACGAGCGGACGCCCGCGACCGCTCACTACC	19665
QY	333	His Pro Asp Ala Ala Phe Leu Phe Ala Asp Arg Ala Ile Ala Leu Asp His Gln Glu Gly	352
Db	19666	GATCCCGAGCGCGCTGCTGCTTCCTGCGACGGGCGCTTCGCTGGAGACCAACCGCACCGCG	19725
QY	353	Cys Cys Tyr Leu Leu Ala Leu Asp Arg Arg Gly His Asp Asp Gly Ala Arg Ala Trp Leu	372
Db	19726	ACCACCTACCGTGGCGGCGGTGGAGAGACGACGCGGACCGGACGAGCGACGCGCTGCTGCTC	19785
QY	373	Arg Glu Thr Ala Glu Thr Leu Thr Gly Leu Ala Val Arg Ala Pro Ala Glu Pro Thr Pro	392
Db	19786	GCGCGCGGCTCCGCGACCTCGAGGCGCCATCGCGCGCGG--GAGCCCGGAGCGG	19836
QY	393	Ala Met Val Phe Gly Ile Pro Glu Ala --Ala Ala Gly Phe Gly Pro Leu Ala Arg Ala	411
Db	19837	TGC-----CCCGAGGCGCGCGTGTGACGACGCGGTCCGGTG--GAGCTG	19878
QY	412	Arg His Asp Yasp Asp Ala Tyr Leu Yasp Arg Ile Asp Glu Cys Leu Yasp Glu Leu Arg Asn	431
Db	19879	CGCCACGACCGCGGACGCGCTACCTGAGACTATGACGCTCTGCCACGACGAGATAGCGCGC	19938
QY	432	Gly Glu Ser Tyr Glu Ile Cys Leu Thr Asn Met Val Thr Ala Pro Thr Glu Ala Thr Ala	451

```

Db      19939  GGGAGAGACCTACGAGGCTCTCCCTGACCAACATGGCCGAGCGGACACGACCTCACCCCG
OY      452     LeuProLeuYrSerAlaLeuAGAlaIleSerProValProThrGlyAlaLeuLeuGlu 471
Db      19999  TGGGGCGGCTACCCCGCGCTGCCCGGGTAGACCCCGCCCTGCCCGCTTCCCTGAGAC
OY      472     PheProGluLeuSerValLeuSerAlaSerProGluArgPheLeuThrIleGlyAlaAsp 491
Db      20059  TTTCGGCCCATGGCCGCTCAGACAGATCTCCGAGCGGTTCTTGCGCATCGACCGGCAC
OY      492     GlyGlyValGluSerIleProIleGlyThrArgProArgGlyGlyThrAlaGluGlu 511
Db      20119  GGGGGGATGAGTCCAGCCGATCAAGGGGACCGGGCCACCGGGCCACCCCGCAGGAG
OY      512     AspGluArgLeuArgAlaAspLeuAlaGlyArgGluLeuAspArgAlaGluAsnLeuMet 531
Db      20179  GACCCCGCGCTCGACGCGCCCTGGCCACTCGGAGAAAGACCGCGCCAGAACTGATG
OY      532     IleValAspLeuValArgAsnAspLeuAsnSerValCysAlaIleGlySerValHisVal 551
Db      20239  ATCGTCGACCTGGTCCGCCACGACCTGGGGCGGTGGCGCGAGTCCGGTCCGTCGCGC
OY      552     ProArgLeuPheGluValGluThrTrpAlaProValHisGluLeuValSerThrIleArg 571
Db      20299  GACCCGGTGTTCAGGTGCGAGAGCTACGCGACCGTCCACAGCTGTGTCAGTACCGCACG
OY      572     GlyArgLeuArgProGlyThrSerThrAlaAlaCysValArgAlaAlaPheProGlyGly 591
Db      20359  GCGGCGCTGCGCCGAGACACAGACCCGGTGGCGGGCGTCCGGCGGCTTCCCGCGGGA
OY      592     SerMetThrGlyAlaProGlyLysArgThrMetGluIleIleAspArgLeuGluGluGly 611
Db      20419  TCGATGACCGGGGGCGCGAAGATCCGACCATGCAATCATGACCGCGTGGAAAGCGGG
OY      612     ProArgGlyValIleSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeu 631
Db      20479  CCGGCGGCTGTCACTCGGGCGGCATCGGCTACTCTCCCTCACCGCGCGGTGAGCTCG
OY      632     SerIleValIleArgThrIleValLeuAlaAspGlyAlaGluPheGlyValGlyGly 651
Db      20539  TCCATCGTATCCGACGAGTGTCTGAGCGGAGCGGTGCGCTCAGCGCTCGCGGCG
OY      652     AlaIleValSerLeuSerAspGlnGluGluPheThrGluThrValValysAlaArg 671
Db      20599  GCGGTCAATCCGCGCTCCGACCGGGCGACAGAGTTGAGAGAGACGGCGGTCAAGCGCGC
OY      672     AlaMetValThrAlaLeuAsp 678
Db      20659  CCGCTGCTGCGTCTCTCCGAC 20679

RESULT 5
US-10-156-761-1167
; Sequence 1167, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMDRA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156, 761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1167

```











```

Db 1469344 GCTCAGTTTGTCTGCTCGGCGCAAGCCGACCGCCCGCCGCGCCAGC----- 1469296
Qy 397 YLLEPGLUALAAlaAaglyPheglyProleuAlaArgAlaarghisaplysaPal 417
Db 1469295 -GCACCCCGCTCGATCGCGGAGCCGCGTGGTGTGCGC-----GACCGTAGCAG 1469246
Qy 417 aTYrLeuLYaArgLYleaspGLuCYleuLYeGLuLlaArgAsnGLyGLuSerTYrGLuL 437
Db 1469245 CTACCTCGCGGACATCGGAGCCTGCAAGCGGAACTAGAGACGCGACCGATTACGAGAT 1469186
Qy 437 eCYLeuThraSmeValThrAlaProThnGLuAlaThrAlaLeuProleuTYrSerAl 457
Db 1469185 CTGTCTACCAACAGCCGCGGCGAGCGTCCCGCCCGCTAGACAGCCCTTACCGGGT 1469126
Qy 457 aLeuArgAlaIleSerProValProTYrGLyAlaLeuLeuGLuPheProGLuLeuSerVa 477
Db 1469125 ACTGCGCTGCTCGACCGCGCGCGCTACCGCGCTTACTGTGGGTTCGTGACCTCGATGT 1469066
Qy 477 lLeuSerAlaSerProGLuArgPheLeuThrLleGLyAlaAspGLyGLyValGLuSerLY 497
Db 1469065 GCGCGGCTGCTCGCCCGGAGGTTTCTGCGATCACCGCGACGCGCTCGCGAGCGCAG 1469006
Qy 497 sProlleYsGLYThraRgPProArgGLYGLYThraLagLugLusapGLuArgLeuArgAl 517
Db 1469005 ACCCGTAAAGGACCGCGCCCGCGCGCGAGCGCGGAGAGAGACCGCGCTACCGGA 1468946
Qy 517 aAspLeuAlaGLYArgGLuLYsaAspArgAlaGLuAsnLeuMeLleValAspLeuValAr 537
Db 1468945 CGCGCTACCAACGAGACAAAGACCGCGCGAGAACCTGGTGTGTCGTGACCTGCTCGC 1468886
Qy 537 gAenAspLeuAsnSerValCYsaAlaIleGLYSerValHisValProArgLeuPheGLuVa 557
Db 1468885 CAACGACCTGGCGCGGCTCGCGAGCCGAGACGCTAGAGTCAACCGCTGATGCGGAC 1468826
Qy 557 lGLuThraYAlaProValHisGLuLeuValSerThrLleArgLYaRgLeuArgProGL 577
Db 1468825 CGAGACCTGCGCCACCGCTGACCACTGTGTCTCACCGTGAAGGGGCGCGCGAGGG 1468766
Qy 577 YThrSerThraAlaAlaCYsaValArgAlaAlaPheProGLYGLYSerMeThrGLyAlaPr 597
Db 1468765 CATCGAGCGGTGAGCTCGGATGCTGTCTTCCCGAGAGTTCCGTGACCGCGGCGCC 1468706
Qy 597 oLYeLYsaRgThrMeGLuLleLleAspArgLeuGLuGLYProArgLYaLYaLYse 617
Db 1468705 GAAACCTCGCACATGAGATCATCTGCTGAGACCGAGCCCGCGGCTGTACTC 1468646
Qy 617 rGLYAlaLeuGLYThrPheAlaLeuSerGLYAlaAlaAspLeuSerLleValLleArgTh 637
Db 1468645 CGAGGCACTCGGCTACTTCGAGTGCAGTGGCGCGCGCCCTCGCCATCGCCAGCGCAC 1468586
Qy 637 rLleValLeuAlaAspGLYGLuAlaGLuPheGLYValGLYGLYAlaLleValSerLeuSe 657
Db 1468585 GCGCGTGTTCACCGCGGAGATGCATCTCGCGCGCGCGGTGCGATCTCTCGGCTC 1468526
Qy 657 rAspGLuGLuGLuPheThrGLuThrValValLYsaArgAlaMeValThrAlaLe 677
Db 1468525 CGATCCGCTCGCGAGTACACGAGTGTCTGMAAGCGCGCACCGATGCGCGCC-- 1468468
Qy 677 uAspGLYSerAlaValAlaGLYAlaArg 686
Db 1468467 -CACCGGACCGGATCGCGGCTGCGG 1468441

```

```

; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, MAKINO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1111
; LENGTH: 1860
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-1111

Alignment Scores:
Pred. No.: 4,25e-138 Length: 1860
Score: 1467.00 Matches: 328
Percent Similarity: 60.06% Conservative: 78
Best Local Similarity: 48.52% Mismatches: 204
Query Match: 41.62% Indels: 66
DB: Gaps: 13

US-10-089-514-2 (1-686) x US-09-738-626-1111 (1-1860)
Qy 1 MetArgThrLeuLeuLleAspAntYrAspSerPheThrHisAsnLeuPheGLuTYrLle 20
Db 1 ATCGCGGTTTATATTATGATATATTGATCTTTCACGTTTATCTCGCACCTATGTC 60
Qy 21 GLYGLuAlaThrGLYGLuLProProValValValProAsnAspAlaAspTrpSerArgLeu 40
Db 61 GAAGAGGTTTACGGGTGAGGACACCTGTGTGTCTTAATCATCAAGAAATAGATGATG 120
Qy 41 ProValGLuAspPheAspAlaLleValLleValSerProGLYProGLYSerProAspArgL 60
Db 121 CTT-----TTCAACCGCGATCTCTCACTGCGCGCGGCGCACCGCGCGTTCG 171
Qy 61 ArgAspPheGLYlLeSerArgArgAlaLleThraSperGLYLeuProValLleGLYVal 80
Db 172 GCTGATTTTGTATCTGTGACGCGCTATGAGCGTGCACGCGTTCGATTTTGGGTGTC 231
Qy 81 CYLeuGLYHisGLuGLYlLleAlaGLuLeuPheGLYGLYThraValGLYLeuAlaProGL 100
Db 232 TGTTTAGGCACACAGGACATTTGGTGGCTTATGCGGAGATGATTTGGCGCCAG 291
Qy 101 ProMeHisGLYArgValSerGLuValArgHisThrChLYGLuAspValPheArgGLYLeu 120
Db 292 CCGGTCCAGGTGAGGTTTCGAGATCCACCATGATGATTCAGGTTTATTTGACGATC 351
Qy 121 ProSerProPheThraAlaValArgTYrHisSerLeuAlaAlaThraAspLeuProAspGL 140
Db 352 CTGGAACGTTTGAAGCGGCTGCTTATCATCTGATGTGGGCAACCGCGTTGCGGAGTCA 411
Qy 141 LeuGLuProLeuAlaTrpSerAspArgLYaValMeGLYLeuArgHisArgGLYLeu 160
Db 412 TTGAAGCTAACAGCTACAGCATGATGTTGATGATCATGATTCGACATGAAAGTCTT 471
Qy 161 ProLeuTrGLYValGLuPheHisPProGLuSerLleGLYSerAspPheGLYArgGLuLle 180
Db 472 CGGACGTGGGTGTGGAATTTCACTCCGAATGATTTGGGCAATTCGCGCATCGATC 531
Qy 181 MetAlaAsnPheArgAspLeuAlaLeuAlaHisHisArgAlaArgArgHisGLYAlaAsp 200
Db 532 ATTGAAGACTTCTTATTTATGCGCGCACATATCGC----- 567

```



QY	201	SerProTyrGluLeuHisValArgArgValAspValLeuProAspAlaGluGluValArg	220
Db	568	-----TGGCAACTCAGCAGAAAACTATTCCGCTCACCGCTTGAATTCAGACGCGCTTTT	621
QY	221	ArgGlyCysLeuProGlyGluGlyThrThrPheTrpLeuAspSerSerValLeuGlu	240
Db	622	GAACATTCTTTGGCCATTCTCCCATGCTTTTGGCTCGATGAT-----	666
QY	241	GlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTyrLeuThr	260
Db	667	---GCCCAAGGAACACGCTATCTTGATGTCACAGCGTCTCTCGACGCAAAAAACC	723
QY	261	TyrArgValAlaAspGlyValValSerValArgGlySerAspGlyThrThrArgThr	280
Db	724	CATAATGTCGGGAGGGG-----	741
QY	281	ArgArgProPhePheAsnTyrLeuGluGluGluLeuGluArgArgValProValAla	300
Db	742	-----GATTTCTTCACTGGCTAAAGAGGATCTCGCCGCAC-----TCAGTTGCG	789
QY	301	ProGluLeuProPheGluPheAsnLeuGlyTyrValGlyTyrLeuGlyTyrGluLeuLys	320
Db	790	CCCGGTCAA-----GTTTCGTCTTGCGTGGGTGGTTACGTTAGAGCTTAA	843
QY	321	AlaGluThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe	340
Db	844	GCGGAAGCTGGCGCACGCGGCTGGCGACACTTCGATCTTCGGGATCGCCACCTCAATTTT	903
QY	341	AlaAspArgAlaIleAlaLeuAspHisGluGluGlyCysCysTyrLeuLeuAlaLeuAsp	360
Db	904	GCCGATCGCGGCATCGACGTGGAAATCGGATCAG-----GTTCCGTTCTGACGTTG--	954
QY	361	ArgArgGlyHisAspAspGlyAlaArgAlaTrpLeuArgGluThrAlaGluTyrLeuThr	380
Db	955	-----GGGAGACAGAGAC-----GAGTGGTTGAAGAAACATCAAGACGCTGAT	999
QY	381	GlyLeu---AlaValArgAlaProAlaGluProThrProAlaMetValPheGlyTyrLeuPro	399
Db	1000	AATCTTTCGCGCCCGCGGATTCCTGCTGC-----	1029
QY	400	GluAlaIleAlaGluPheGlyProLeuAla---ArgAlaArgHisAspGlyAspAlaTyr	418
Db	1030	-----GACACCTCGCTTTCGACGGTTCGAGATTCCAAAAGATGAGTAT	1071
QY	419	LeuLysArgGlyLeuAspGluCysLeuLysGluIleArgAsnGlyGlySerTyrGluIleCys	438
Db	1072	CTCGCAAAATTCGACAGCCACGAGCGGTATTACTCGCGCGAATCGTAAGAATCTGC	1131
QY	439	LeuThrAsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeu	458
Db	1132	CTGACCACAAAACTTCAGGGCACCACTATGCGCCCTCTTGCGTCCCTACTTAGCACTG	1191
QY	459	ArgAlaIleSerProValProTyrGlyAlaLeuLeuGluPheProGluLeuSerValLeu	478
Db	1192	CGTGGGCGCAATCCACCGCAATGTGTCGATCTTCAGCTGGGGGAAATCACTTATTTTG	1251
QY	479	SerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGluSerTyrPro	498
Db	1252	AGTTCCTCGCGGAGCGGTTCATCACTTATTCGGAGGGGTATGTGGATCAAGGCC	1311
QY	499	IleLysGlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuArgAlaAsp	518
Db	1312	ATTAAGGACACGAGCCCGCTGGCGGAACAGCGCAAGAAACCAAGAAATCATTTGCTGAG	1371
QY	519	LeuAlaGlyArgGluLysAspArgAlaGluAsnMetIleValAspLeuValArgAsn	538
Db	1372	CTGGCGCAATATCTTAAGAGTCGTGCAAAAACTTGATGATCGTGGAATTTGTCCGCAAC	1431
QY	539	AspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGlu	558
Db	1432	GACTTAGCCCGCGGCGCTTGTGCCACCACTGAATTAACATCAAGCTTTTGGAGCTGCA	1491
QY	559	ThrTyrAlaProValHisGluLeuValSerThrIleArgGlyArgLeuArgProGlyThr	578

```

Db      1492  ACCTAGCGCACAGCTCCACCAACTTGTCAGACCCGCTCTCTGACAGATTGGGCGCA--CGC 1548
Oy      579   SerThrAlaAlaCyValArgAlaAlaPheProGlyGlySerMetCThrGlyAlaProLys 598
Db      1549  AGTCGGATTGAGTCGCTCGCGGACGACATCTCCCGGTGGTTGATGACTGGTGCCCAAG 1608
Oy      599   LysArgThrMetGluLeuLeuIleAspArgLeuGluGluGlyProArgGlyValTyrSerGly 618
Db      1609  CTGGCGACCATGAGATCATCATGACTGAGACCTGGAGCACCTCCTCCGGGTATTATCTCAGGT 1668
Oy      619  AlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIle 638
Db      1699  GGCCTTGGGATATTTTTCCTTCGACGGCGCAGATTATCTCTCCACAGGATCAGAACTCTC 1728
Oy      639  ValIleuAlaAspGlyGlnAlaGluPheGlyValGlyGlyAlaIleValSerLeuSerAsp 658
Db      1729  GTCATCCAGAACATCATCAGTGGAGTATGAGAGTGGCGCGGTGCATCTTTCCTCTGTCTGAT 1788
Oy      659  GlngGluGluGluPheThrGluThrValValIysAlaArgAlaMetVal 674
Db      1789  CCGGAGCGCTGAGTGGAGAAATCCGCGTTAAATCATCAGGCGCTCTGCTG 1836

RESULT 8
US-10-494-675-25
/ Sequence 25, Application US/10494675
/ Publication No. US20050019877A1
GENERAL INFORMATION:
/ APPLICANT: Zelder, Oskar
/ APPLICANT: Pompejus, Markus
/ APPLICANT: Schroder, Hartwig
/ APPLICANT: Krieger, Burkhard
/ APPLICANT: Klopriegse, Corina
/ APPLICANT: Haberhauser, Gregor
TITLE OF INVENTION: Genes coding for metabolic pathway proteins
FILE REFERENCE: BGI-163US
CURRENT APPLICATION NUMBER: US/10/494,675
CURRENT FILING DATE: 2004-05-04
PRIOR APPLICATION NUMBER: PCT/EP02/12141
PRIOR FILING DATE: 2002-10-31
PRIOR APPLICATION NUMBER: DE 101 54 292.1
PRIOR FILING DATE: 2001-11-05
NUMBER OF SEQ ID NOS: 164
SEQ ID NO 25
LENGTH: 2005
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (101)..(1975)
/ OTHER INFORMATION: RXA005079
US-10-494-675-25

Alignment Scores:
Pred. No.: 4,69e-138 Length: 2005
Score: 1487.00 Matches: 328
Percent Similarity: 60.06% Conservative: 78
Best Local Similarity: 48.52% Mismatches: 204
Query Match: 41.62% Indels: 66
DB: 21 Gaps: 13

US-10-089-514-2 (1-686) x US-10-494-675-25 (1-2005)
Oy      1  MetArgThrLeuLeuIleAspAsnTyrAspSerPheThrIleAsnLeuPheGlnTyrIle 20
Db      116  ATGGCGGCTTTAAATATTGATTAATTAATGATTTCTTTCAGCTTAAATCTCGCACCTATGTTG 175
Oy      21  GlyGluAlaThrGlyGlnProProValValValProAsnAspAlaAspTyrPseArgLeu 40
Db      176  GAAAGAGGTACGGGTACAGCACCTGTGTGTGTGCTTAATGATCAAGAAATAGATGAGATG 235
Oy      41  ProValGlnAspPheAspAlaIleValValSerProGlyProGlySerProAspArgGlu 60

```



Db 236 CTT-----TTGACGCCGTCATCTCTACCTGCGCCGGCCGACGCCGGCTTCCG 286  
 Qy ArgAspPheGlyIleSerArgAlaIleThrAspSerGlyLeuProValLeuGlyVal 80  
 Db 287 GCGATTTTGGTATCTGTGACGGGGCTTATGAGGTGACCGCTTCCGATTTTGGGTGTG 346  
 Qy CysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaProGlu 100  
 Db 347 TGTATTAGGCCACACGAGGATTCGCTGGCCATATGCGGATGATGTTGATTGGCCCGCAGG 406  
 Qy 101 PromethisGlyArgValSerGlyValArgHisThrGlyGlyLeuValPheArgGlyLeu 120  
 Db 407 CCGGATCCACGGTGAAGTTTGCAGATCACTCATGATGATTCAGGTTTATTTGACGGGATC 466  
 Qy 121 ProSerProPheThrAlaValArgTyrHisSerLeuAlaIleThrAspLeuProAspGlu 140  
 Db 467 CTTGAACGTTTGAAGCGGCTGCTTATCATCTCGATGGCCACCCGCTTCCGGAATCA 526  
 Qy 141 LeuGlnProLeuAlaIleTyrSerAspAspGlyValValMetGlyLeuArgHisArgGlyLys 160  
 Db 527 TTGAAAGCTACAGCTACAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 586  
 Qy 161 ProLeuTyrGlyValGlnPheHisProGlySerIleGlySerAspPheGlyArgGlyIle 180  
 Db 587 CCGCAGTGGGGTGTGCAATTTTCATCCGGAATCTATTGGTGCAATTCGGCCATCAGATC 646  
 Qy 181 MetAlaAsnPheArgAspLeuAlaLeuAlaHisIleHisArgAlaArgAlaGlyValAsp 200  
 Db 647 ATTGAAGACTCTCTTAATTGACCGGACATATGCG-----682  
 Qy 201 SerProTyrGlnLeuHisValArgArgValAspValLeuProAspAlaGlnGluValArg 220  
 Db 683 -----TGGCAACTCAGCGAGAAACATATTCCTCGTCAAGCTTATGACGACGGCTTTT 736  
 Qy 221 ArgGlyCysLeuProGlyGlnGlyThrThrPheTyrLeuAspSerSerSerValLeuGlu 240  
 Db 737 GAAACATTTCTTGGCCATCTCCATCGCTTTTGGCTCGATGAT-----781  
 Qy 241 GlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGlyTyrLeuThr 260  
 Db 782 ---GCCAAGGAAACACAGCTATCTTGTGATGCCAGCGTCTCTCGGACACCAAAAC 838  
 Qy 261 TyrArgValAlaAspGlyValIleValSerValArgGlySerAspGlyThrThrArgThr 280  
 Db 839 CATATATCCGCGCAGCGG-----856  
 Qy 281 ArgArgProPhePheAsnTyrLeuGlnGlnLeuGlnArgArgValProValAla 300  
 Db 857 -----GATTTCTTCACTGCTTAAGAGGATCTGCGCGCAAC-----TCAGTTTGGC 904  
 Qy 301 ProGlnLeuProPheGlnPheAsnLeuGlyTyrValGlyTyrLeuGlyTyrGlyLeuLys 320  
 Db 905 CCGCGTCA-----GGTTTCTGCTTGGCTGGGCTTGTATGAGCTTAA 958  
 Qy 321 AlaGlnThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe 340  
 Db 959 GCGGAACCTGCGGACCGGCTGCGGACACTTCCAGTCTTCCGATGCGGACCTCATTTT 1018  
 Qy 341 AlaAspArgAlaIleAlaLeuAspHisGlnGlnGlyCysArgTyrLeuLeuAlaLeuAsp 360  
 Db 1019 GCGGATGCGGCAATCGGAGTGAATCGGATCAG-----GTTCGGTTGCTGGCTTGG---1069  
 Qy 361 ArgArgGlyHisAspAspArgGlyAlaArgAlaTyrLeuArgGlyThrAlaGlnThrLeuThr 380  
 Db 1070 -----GGGAGGACGAGAC-----GAGTGGTTTGAAGAAACCATCAAGAGCTGCAT 1114  
 Qy 381 GlyLeu-----AlaValArgAlaProAlaGlnProThrProAlaMetValPheGlyIlePro 399  
 Db 1115 AATTTTGTGCGCCCGGGAATACGCTGTC-----1144  
 Qy 400 GluAlaAlaAlaGlyPheGlyProLeuAla-----ArgAlaArgHisAspLysAspAlaTyr 418  
 Db 1145 -----GGAACCTCTGCTTGTGAGTTGAGATTCGAATCCAAAGATGAGTAT 1186

Qy 419 LeuLysArgIleAspGlyCysLeuLysGlnIleArgAsnGlyGlnSerTyrGlnIleCys 438  
 Db 1187 CTGACAAATTTGGCAGAGCCGACGAGCTGATTAATCTCGGCGCAATCTGTAATATGTC 1246  
 Qy 439 LeuThrAsnMetValThrAlaProThrGlnAlaThrAlaLeuProLeuTyrSerAlaLeu 458  
 Db 1247 CTGACCAAAATCTGACGAGGACCACTGATGAGCCCTCTGCTGCTGCTATTCAGCACTG 1306  
 Qy 459 ArgAlaIleSerProValProTyrGlyAlaLeuLeuGlnPheProGlnLeuSerValLeu 478  
 Db 1307 CGTGGGGCCAAATCCACCGCATATGCGCTATCTTACGCTGGGAGATACCTTAATTTTG 1366  
 Qy 479 SerAlaSerProGlnLysPheLeuThrIleGlyValAspGlyGlyValGlnSerLysPro 498  
 Db 1367 AGTTCTCCCGGAGGCGTTTATCATCATGATGATTTGCGACAGGATGATGATGATGATG 1426  
 Qy 499 IleLysGlyThrArgProArgGlyGlyThrAlaGlnGlnGlnGlnGlnGlnGlnGlnGln 518  
 Db 1427 ATTAAAGGACCAAGCGCGGTGGGGAACAGCGCAAGAGCAAGAAATCATTTGCTGACG 1486  
 Qy 519 LeuAlaGlyArgGlyLysAspArgAlaGlnLeuMetIleValAspLeuValArgAsn 538  
 Db 1487 CTGCGAGTAAATCTTAAGATCGTGCAAAACTTGATGATCGTGATTTGGTCCCGCAAC 1546  
 Qy 539 AspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGlyValGln 558  
 Db 1547 GACTTACCCCGCGCGCTTGGCCACACAGATTAAACATCCAAAGCTTTTGCAGTGCAG 1606  
 Qy 559 ThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyThr 578  
 Db 1607 ACCTTACGCAAGTCCACCACTTGTACAGCACCGCTCTGCGAGATTTGGGCGCA---CGC 1663  
 Qy 579 SerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLys 598  
 Db 1664 AGTCCAGTGAAGTGCCTGCGCGGACGATTCGCGGATTCGATGATGATGATGATGATG 1723  
 Qy 599 LysArgThrMetGlnIleIleAspArgLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 618  
 Db 1724 CTGCGACCAATGAGATCATCATGATGATGATGATGATGATGATGATGATGATGATG 1783  
 Qy 619 AlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIle 638  
 Db 1784 GCGTTGGATATTTTCCCTCCAGCGGCGAGTGAATCTCTCCATGATGATCAGAACTCTC 1843  
 Qy 639 ValLeuAlaAspGlyGlnAlaGlnPheGlyValGlyGlyAlaIleValSerLeuSerAsp 658  
 Db 1844 GTCATCCAGAAATCATCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1903  
 Qy 659 GlnGlnGlnGlnPheThrGlnThrValValIleValAlaArgAlaMetVal 674  
 Db 1904 CCGAGGCTGAGTGAAGGAGAAATCCCGTTAAATCAACGCGCTCTGCTG 1951

RESULT 9  
 US-09-738-626-1  
 ; Sequence 1, Application US/09738626  
 ; Publication No. US20020197605A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: NAKAGAWA, SATOSHI  
 ; APPLICANT: MIZOGUCHI, HIROSHI  
 ; APPLICANT: AMDO, SEIKO  
 ; APPLICANT: HAYASHI, MIKIRO  
 ; APPLICANT: OCHIAI, KEIKO  
 ; APPLICANT: YOKOI, HARUHIKO  
 ; APPLICANT: TATEISHI, NAKO  
 ; APPLICANT: SENOH, AKIHIRO  
 ; APPLICANT: IKEDA, MASATO  
 ; APPLICANT: OZAKI, AKIO  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCES: 249-125  
 ; CURRENT APPLICATION NUMBER: US/09/738,626  
 ; CURRENT FILING DATE: 2000-12-18  
 ; PRIOR APPLICATION NUMBER: JP 99/377484











```
QY 455 TYSerAlaLeuArgAlaIleSerProValProTyrGlyAlaLeuLeuGluPheProGlu 474
Db 1318 TACAAACACTTGGCCAGGAAAGTCCCGTCAGTACTCTCTTTTGGCTTTAGATGAT 1377
QY 475 LeuSerValLeuSerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyVal 494
Db 1378 TTTCAGATTCTTTCCTCTTCATGAGCGGTTTATTAAGTCCACAAGAGCCGGGTGTG 1437
QY 495 GluSerProIleLeuGlyThrArgProArgGlyGlyThrAlaGluGluAspGluArg 514
Db 1438 ACCACTAAGCCATCAAGGGGACGTCAGACGGGGACAGATACCAAGAGACCAAGAC 1497
QY 515 LeuArgAlaAspLeuAlaGlyArgGluYAspArgAlaGluAsnLeuMetIleValAsp 534
Db 1498 TTGATGAAAGGCTCCGACGTAAAGAAACCAAGGACGAAACCTGATGATGTCAG 1557
QY 535 LeuValArgAsnAspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeu 554
Db 1558 CTCTTCGCAATGATCTGGGCGGTTTTTGTGAATCGGGTGGGTGAGTCCCAAACTG 1617
QY 555 PheGluValGluThrTyrAlaProValHisGluLeuValSerThrIleArgGlyArgLeu 574
Db 1618 ATGATGATGAGACCTTACCTCCACCTCCAGCTGGTTACACGGTTCCAGCGGTC 1677
QY 575 ArgProGlyThrSerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThr 594
Db 1678 AAGAAGACCTAGATGTTGTTGAAGTTAGAAAACACCTCCGGGCGGTCATGACT 1737
QY 595 GlyAlaProGlyAlaArgThrMetGluIleIleAspArgLeuGluGluProArgGly 614
Db 1738 GGGGCCCTTAAAAAAGAACCCCGAATATTGATGACTTAGAACAGATGCCAAGAGG 1797
QY 615 ValTyrSerGlyAlaLeuGluYTPheAlaLeuSerGlyAlaAlaAspLeuSerIleVal 634
Db 1798 ATGATATCTGGGCTATGCTTTTAAAGCAATTTCCACTAATGCTTCAACATCGTC 1857
QY 635 IleArgThrIleValLeuAlaAspGlyGlnAlaGluPheGlyValGlyAlaIleVal 654
Db 1858 ATCCGACCAAGCTGTGGTGAAGACAGACAGCAAGCATGCGCTGGGGGGTGGATTTGTC 1917
QY 655 SerLeuSerAspGlnGluGluGluPheThrGluThrValValIleAlaArgAlaMetVal 674
Db 1918 ATGCTATCTGATCTCGAAGAGAGATTGATGAAGTTTAAAGCTAAGGGGCGCTTG 1977
QY 675 ThrAlaLeu 677
Db 1978 TCCGCTTG 1986

RESULT 11
US-10-501-282-6651/C
; Sequence 6651, Application US/10501282
; Publication No. US20050203280A1
; GENERAL INFORMATION:
; APPLICANT: MCMICHAEL, JOHN CALHOUN
; APPLICANT: ZAGURSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARRISH
; APPLICANT: FLETCHER, LEAH DIANE
; TITLE OF INVENTION: ALDOTOXICUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
; TITLE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF
; FILE REFERENCE: AM100780 I2
; CURRENT APPLICATION NUMBER: US/10/501,282
; PRIOR FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6651
; LENGTH: 1754382
; TYPE: DNA
```

```
; ORGANISM: Altolococcus ostitidis
US-10-501-282-6651

Alignment Scores:
Pred. No.: 1,41e-120 Length: 1754382
Score: 1350.00 Matches: 236
Percent Similarity: 58.71% Conservative: 105
Best Local Similarity: 43.34% Mismatches: 248
Query Match: 37.78% Indels: 34
DB: 22 Gaps: 12

US-10-089-514-2 (1-686) x US-10-501-282-6651 (1-1754382)

QY 1 MetArgThrLeuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGlnTyrIle 20
Db 7807 ATGAAGTCACTCATTTTCATTAACATTAATGATTTCTTATCTTACACTTATACAGCTTATT 7748
QY 21 GlyAlaAlaThrGlyGlnProProValAlaValProAsnAsp---AlaAspTrpSerArg 39
Db 7747 GGGAAAGTCAACAGGAAAGGCCCATGCTGATATAAAATGACCAATGACCTTACCAAGAA 7688
QY 40 LeuProValGluAspPheAspAlaIleValValSerProGlyProGlySerProAspArg 59
Db 7687 CTATTGATCTCGACTTGTATGATGATCATTTATTTCAACCGGGCCCGGTGTCGGACCGG 7628
QY 60 GluArgAspPheGlyIleLeuSerArgAlaIleThrAspSerGlyLeuProValLeuGly 79
Db 7627 GACAAAGACTTTCGGCTTGGTCGGCAGATTAATGAAAAGCTTGACAAAGCTTATCTTTGGC 7568
QY 80 ValCysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaPro 99
Db 7567 ATTGCTGGGCGACAGGCGCATTTATTTTAACGGTGGGCACTTATGGGGCGCTGAT 7508
QY 100 GluProMetHisGlyArgValSerGluValArgHisThrGlyGluAspValPheArgGly 119
Db 7507 ATCCCATCATGTCGCGGAGATTCGCTCCATTAATGAGGAAATATCTTTGACGCGC 7448
QY 120 LeuProSerProPheThrAlaValArgTyrHisSerLeuAlaAlaThrAspLeuProAsp 139
Db 7447 ATAGACCAAGGCTTTGAACTCAACGCTACCTCATTTGTTGGAAGAC-----AAG 7394
QY 140 GluLeuGluProLeu-----AlaTrpSerAspAspGlyValValMetGlyLeuArg 156
Db 7393 GAATTAGAAATATTCATTCATGATGCCAACAACGATGATGAAATTGTCATGGCCCTGTC 7334
QY 157 HisArgGluLeuProLeuTrpGlyValGlnPheHisProGluSerIleGlySerAspPhe 176
Db 7333 CACAAGACCAAGCCCATCTGATGGGTCATTCACACCCGAATCCATTTGCCACCCAGTAT 7274
QY 177 GlyArgGluIleMetAlaAsnPheArgAspLeuAlaLeuAlaHisArgAlaArgArg 196
Db 7273 GGGAAAAAATATGATGAAAACTTTATGGCTTGTGCGACGACTACTAT----- 7226
QY 197 HisGlyAlaAspSerProTyrGluLeuHisValArgArgValAspValLeuProAspAla 216
Db 7225 -----AACCAATCCAGGCTTTATTTGAAAAAGTTCCGGGAACCTAGGACACAC 7178
QY 217 GluGluValArgArgGlyCysLeuProGlyGluGlyThrPheThrLeuAspSerSer 236
Db 7177 CAAGACCTTACGATCAATTTGCGCAAAATTTGATGACCAAGTACTTTGGCTCAGCTCCAGC 7118
QY 237 SerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAla 256
Db 7117 AAGGTGAGCCTGGCTGTGCAGGTTTCCATCTTTGGATGGCTGGACCAAAAGCGGGC 7058
QY 257 GluTyrLeuThrTyrArgValAlaAspGlyValValSerValArgGlySerAspGlyThr 276
Db 7057 CACACTTAAATATGATGATCCACCAAAAAGAGTGAAGAAAAACAAGATGCGCAGTCCGA 6998
QY 277 ThrThrArgThr---ArgArgProPhePheAsnTyrLeuGluGluGluLeuGluArgArg 295
Db 6997 GCAGTTGAAGGCTCCAGACTGACATCTTTCTTACTTA-----AAAGCCAAACGACCA 6944
```



QY	296	ArgValAProValAlAProGluLeuProPheGluPheAsnLeuGlyTYrValGlyTYrLeu	315
Db	6943	AAATGGCTTATGAGCCGGGCTTCCTTTGACTTTCAGCTTAGGGCTAATATGGCTAATATC	6884
QY	316	GlyTYrGluLeuLysAlaGluThrThrgly--AspProAlaHisArgSerProHisPro	334
Db	6883	GGCTATGAGGTCAAAAAGATGACGTTCAAGTCCAAACAAAGACCACTACCCATCG	6824
QY	335	AspAlaAlaPheLeuPheAlaAspArgAlaIleAlaLeuAspHisGlnGluGlyCysCys	354
Db	6823	GATGGCTCTATTACTTATGTGACCGGGCTTGGTCTATGACCAACAGAAAGAGACCTT	6764
QY	355	TyrLeuLeuAlaLeuAspArgArgGlyHisAspAspGlyAlaArgAlaThrPheArgGlu	374
Db	6783	TACTTCCTCTCTAC-----CAGGATGC-----CAGATTGATTTCAACG	6722
QY	375	ThrAlaGluThrLeuThrglyLeuAlaValArgAlaProAlaGluProThrProAlaMet	394
Db	6721	GTCAAAGACAGGCTCAAG---CAGGCATCAAGTTCAGCAAGAAAGA-----	6677
QY	395	ValPheGlyIleProGluAlaAlaIleGlyPheGlyProLeuAlaArgAlaArgHisAsp	414
Db	6676	-----CCAAAGAGCAGACAGAAAGCTCCACAGCATGGCTTGCTGTCAA---CAC	6632
QY	415	LysAspAlaTYrLeuLysArgIleAspGluCysLeuLysGluIleArgAsnGlyGluSer	434
Db	6631	AAAGCAGGCTTATATAAAGACATTGAAACATCCAAAGCTGATTCMAAGCTGGGGAAAGT	6572
QY	435	TyrGluIleCysLeuThrAsmMetValThrAlaProThrgluAlaThrAlaLeuProLeu	454
Db	6571	TATGAGTTGCTTACCAACCGGCTGGATATTGAGGGTCAGATTGATGAGCTTGCTTAT	6512
QY	455	TyrSerAlaLeuArgAlaIleSerProValProTYrGlyAlaLeuLeuGluPheProGlu	474
Db	6511	TACAAACACTGGCGCAGAGAAAGTCCCGAGTCACTCTGCTTTTGCCTTTAGATGAT	6452
QY	475	LeuSerValIleuSerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyVal	494
Db	6451	TTTCAAGTCTTCTCTCTTCCATGAGGGTTTATTAATGTCGACMAAGACGGGGTGTGTG	6392
QY	495	GluSerLysProIleLysGlyThrxArgProArgGlyValThrAlaGluLysAspGluArg	514
Db	6391	ACCACCTAAGCCCTTAAGGGAGCGTCAAGCGGGGACAGATGACCAAGAAAGACCAAGAC	6332
QY	515	LeuArgAlaAspLeuAlaGlyArgGluLysAspArgAlaGluAsnLeuMetIleValAsp	534
Db	6331	TTGATAGAAAGGCTCCGACGAGAAAGAAACCAAGCAGAAACCTGATATTGTGCAC	6272
QY	535	LeuValAlaGluAspLeuAsnSerValLysAlaIleGlySerValHisValProArgLeu	554
Db	6271	CTCTTCCCAATGATCTGGCGCGTTTGTGAATCCGGGTGGGTGAGAGTCCCAAACCTG	6212
QY	555	PheGluValGluThrTYrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeu	574
Db	6211	ATGGATGTAGAGACCTACTCTCAACCTCCACACGAGCTGTTACACAGGTTTCAGCGCGGCTC	6152
QY	575	ArgProGlyThrSerThrAlaAlaCysValArgAlaAlaPheProGlyIleSerMetThr	594
Db	6151	AAAGAAAGCCTAGAGCTGTGTAAGCTTAAAGAAACACCTTCCCGCGGGGCTCATGACT	6092
QY	595	GlyAlaPLeuLysArgThrMetGluIleIleAspArgLeuGlnGluGlyProArgGly	614
Db	6091	GGGGGCGCTTAAAAAAGAACCTCGAATTTATTGATGACTTAGAAGCAGTTCCACAGAGG	6032
QY	615	ValTYrSerGlyAlaLeuGlyTYrPheAlaLeuSerGlyAlaAlaAspLeuSerIleVal	634
Db	6031	ATCTATTTCTGGACTATTGGGCTTTTATCCCAACATTCCTACTATGACTTCAACATGCTC	5972
QY	635	IleArgThrIleValLeuAlaAspGlyGlnAlaGluPheGlyValGlyValAlaIleVal	654
Db	5971	ATCCGAGCAGAGTGTGTGGAGACAGACAAAGCAGACATCGGCTGGGGGGTGCATTGTC	5912
QY	655	SerLeuSerAspGlnGluGluPheThrgluThrValValLysAlaArgAlaMetVal	674

[illegible]



Db 2525 GAGTGTCTACTCTTTAATCATCATCCATCCAGTAAATGAGTTCAGAGTAGGCGG 2466  
QY 129 TYRHisSerLeuAla-----AlaThrAspLeuProAspGluProLeuAlaIlePro 146  
Db 2465 TATCATCTCACTTGTATATAGAACAGACTCTATCTGAGAGATCTTATATCAATAGCATGG 2406  
QY 147 -----SerAsp----- 148  
Db 2405 ACTGTCTTCGAAAAATGCTCTCATTCCTTGAAGGAGTAAAGCTGATATATAGTACAGT 2346  
QY 148 ----- 148  
Db 2345 ACCCTGTGGGATCATTTGATTAATTATTCGTAACAAACGATGAGTGCAGTACCACT 2286  
QY 149 -----AspGly-----ValIleMet 153  
Db 2285 GATGAAAAATGCCCGACATAAACGATGCAAGTGAAGTGAAGTGAAGGAGTTCCTCATG 2226  
QY 154 GYLeuArgHisArgGluProLeuTrpGlyValGlnPheHisProGluSerIleGly 173  
Db 2225 GGGCTTAGGACTCTACAGAGCCCTCATATGAGTGAAGTTCACCCCGAGAGTGTGGT 2166  
QY 174 SerAspPheGlyArgGluIleMetAlaAsnPheArgAspLeuAla----- 188  
Db 2165 ACTCATTTATGAAAGACAGATATTTCAAAACTTCAGAAAGATAACTGACTTTGGATTA 2106  
QY 189 -----LeuAla 190  
Db 2105 CAGACACCATGCTTCAGGAAAAAAGATCATGTTCCAGATCTCAAGAACTTTGTGTGA 2046  
QY 191 His-----HisArgAlaArg----- 195  
Db 2045 AATGACTTGTATTCATCTGCAAGAGTTGAACCTTTGGGATTCGTGGCCTTGCTGCTT 1986  
QY 196 -----ArgHisGlyAlaAspSerProIlyrGluLeuHisValArgArgValAsp----- 211  
Db 1985 CCAAGCGAAGCAGTGGGGGCAAAATGCTTAAGGTTCGATGAGAAAAAGATGTAACTTC 1926  
QY 212 -----ValLeu-----ProAspAla 216  
Db 1925 CTCATTCGCAATAGTGGCTCTGAAAAATTTCCTGAGTGTTCGCTTTGGCCATCATAGCGT 1866  
QY 217 GlnGluValArgArgGlyCysLeuProGlyGluGlyThrThrPheTrpLeuAspSer 236  
Db 1865 GAAGAT-----ACATTTGGCTGGATGCTCA 1839  
QY 237 SerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAla 256  
Db 1838 TCAGTTGACCGAATAGGGCAGCATTTTCATTCATGGCGGCAAGGGCGGCTTTGG 1779  
QY 257 GlnTrpLeuThrTyraArgValAlaAsp-----GlyValAlaSerVal 270  
Db 1778 AAGCAATATGACATTTTCACTTCGCTCCAGTCAACGACCAATTTGTGAGAGAACTTACTATT 1719  
QY 271 ArgGlySerAspGlyThrThrThrArgThr-----ArgArgProPhePheAsnTrpLeu 288  
Db 1718 CCGAGATGCTTATGTTGACTGTCAAGAACTTCTCAAGATGGTTCCTTGGATTCCTT 1659  
QY 289 GlnGluGlnLeuGluArg--ArgArgValProValAlaProGlu--LeuProPheGlu 306  
Db 1658 GAAAGAGAGATGCAATTCATTCATTAATGAAAGAGATTAAGAGACCTTCATTTGAGC 1599  
QY 307 PheAsnLeuGlyTyraValGlyTyraLeuGlyTyraGluLeuGlyAlaGlu--ThrThrGly 325  
Db 1598 TTCATATGCTGAGATTTGTTGGATTCATAGATATGCTTTAAAGTTAAATGCGATGATCA 1539  
QY 326 AspProAlaHisArgSerProHisProAspAlaAlaPheLeuPheAlaAspArgAlaIle 345  
Db 1538 TCATAATAGTCAAAATCAAGTACCTCGATGATGCTTCTTCTTGTGATTAACATAGTT 1479  
QY 346 AlaLeuAspHisGlnGluGlyCysTyraTrpLeuLeuAlaLeuAsp-----ArgArg 362

Db 1478 GTGGTTGATCAACAACATGCGGATGTATCATTTTATCATTCATGATGAATATTTCTTCT 1419  
QY 363 GYHisAspAspGly-----AlaArgAlaTrpLeuArgGluThrAla 376  
Db 1418 GGTAAATGAGATGGAGATTACCAAAAATCATATACATAGTTATGTTTACCAATAATCTGAG 1359  
QY 377 GlnThrLeuThrGlyLeuAlaValArgAlaPro----- 387  
Db 1358 AAGAGCTTCTCAGATGATGATCCATGGGCCCAAGATTATGATCAATGAAACCTGTGG 1299  
QY 388 -----AlaGluProThrProAlaMetValPheGlyIlePro 399  
Db 1298 ATCAATGGGAACCTATTATCCATATCATTCAGTGTGAATAGCAAAAGATTTGTCTGAG 1239  
QY 400 GlnAlaAlaIleArgIlePheGlyProLeuAlaArgAlaArgHisAspIysAspAlaTrpLeu 419  
Db 1238 AATATCA-----AAAATGAATATATC 1218  
QY 420 LysArgIleAspGluCysLeuLysGluIleArgAsnGlyGluSerTrpGluIleCysLeu 439  
Db 1217 AGAGATGTGCAGAGTGTCTGATTAATAGAGACGAGAAAGCTATGAAATTGTGCTTA 1158  
QY 440 ThrAsnMetValThrAlaProThrGlu--AlaThrAlaLeuProLeuTrpSerAlaLeu 458  
Db 1157 ACTACTCAGATGAAGAAACGAGATTATATGATGCTTTGAACCTTAACCTGAAATTTG 1098  
QY 459 ArgAlaIleSerProValProTrpGlyAlaLeuLeuGluPhe-----ProGluLeuSer 476  
Db 1097 CGAAAAACAATAATCCAGCCCTTATATGACGCTTGCGTTAACTTCTCCGCAAAAAACCTGAT 1038  
QY 477 ValLeuSerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGluSer 496  
Db 1037 AATATGTTGCTCTTCTCTGAA----- 1017  
QY 497 LysProIleLysGlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuArg 516  
Db 1017 ----- 1017  
QY 517 AlaAspLeuAlaGlyArgGluLysAspArgAlaGluAsnLeuMetIleValAspLeuVal 536  
Db 1016 -----AGTAAAAAGATCAAGCTGAGAACTTGATGATGATTCCTGACCTCCGA 972  
QY 537 ArgAsnAspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGlu 556  
Db 971 AGAAACGATCTGTGAAGTGTGTGGAACCTGGGAGTGCATAGTTCCTGCTCATAGAT 912  
QY 557 ValGluThrTyraIleProValHisGlnLeuValSerThrIleArgGlyArgLeuArgPro 576  
Db 911 GTGGATCATATTAACCTGTATACATCCATGATGAATGCCATTCGTGGAACCAAAATGTGG 852  
QY 577 GlnTrpSerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAla 596  
Db 851 GACCTTAAGCCCTGTGATTTGTAAAGCTGTCTTCCGAGAGTTCAATGACCGGGGCC 792  
QY 597 ProLysLysArgThrMetGluIleIleAspArgLeuGluGlyProArgGlyValTrp 616  
Db 791 CCGAAAGTTAATCATGAGATGATCTTGATTCACCTGAAATCAATGCCGAGAGAAATATAC 732  
QY 617 SerGlyAlaLeuGlyTrpPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArg 636  
Db 731 TCAGGATCGTTGATCTTTTCATATTAACAAAGACTTTGACTTGAAATTCGTGATCAGA 672  
QY 637 ThrIleValLeuAlaAspGlyGlnAlaGluPheGlyValGlyValAlaIleValSerLeu 656  
Db 671 ACAAGTTTCTTCACATATGAGAGAGCTTGATTTGGGCGAGGGGCTATTTGTAGCATTTG 612  
QY 657 SerAspGlnGluGluPheThrGluThrValValLysValArgAla 672  
Db 611 TCAGATCCAGAAAGCAGAGTAAATGATGATGCTGTTAAAGCAAAAGCT 564

RESULT 13  
US-10-282-122A-17825  
; Sequence 17825, Application US/10282122A



```

; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykend, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17825
; LENGTH: 2523
; TYPE: DNA
; ORGANISM: Corynebacterium diphtheriae
US-10-282-122A-17825
Alignment Scores:
Pred. No.: 3,94e-96 Length: 2523
Score: 1068.50 Matches: 267
Percent Similarity: 50.65% Conservative: 82
Best Local Similarity: 38.75% Mismatches: 239
Query Match: 29.90% Indels: 101
DB: Gaps: 20
US-10-089-514-2 (1-686) x US-10-282-122A-17825 (1-2523)
QY 4 LeuLeu1leAspAntyYrAspSerPheThrHisAsnLeuPheGlnTyrlleGlyLualA 23
DB 19 TTGGTGTGATATTTCATCTCTTCACTCCACAAACATCGCATCTCCACCGGTGC 78
QY 24 ThrGlyInProProValValVal-----ProAsnAspAlaAspTPSerArg 39
DB 79 GGGGCCCGCACCCACGCGTCAACCAACGATATCCAGGACATTGATTGACCGC 138
QY 40 LeuProValGluAspPheAspAlaIleValValSerProGlyProGlySerProAspArg 59
DB 139 -----TACCAACGAAATAGTCATCTCCACAGGTCCAGGCGCATCCAGTGTG 183
QY 60 GluArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeuGly 79
DB 184 GCAGAAACGTCGCGCATTAAGCGCGGTGGTATTACAAACAGCACAGTCCCAAGTCTGGCG 243

```

```

QY 80 ValCysLeuGlnHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaPro 99
DB 244 GTGTGCTTAGGTATAGCATGCTGATGTGCACAAAGGAGGGGCTTCGTTGATCGGCCCG 303
QY 100 GluProMetHisGlyArgValIserGluValArgHisThrGlyGluAsp---ValPheArg 118
DB 304 GAGGCTGCCACGAGCGGTAGATACCTTAATATTGTGCGCGCATGAGCTTTTGCT 363
QY 119 GlyLeuProSerProPheThrAlaValArgTyHisSerLeuAlaAlaThrAspLeuPro 138
DB 364 GGTCTTACCTCGAAGTTTTCATGTCGCGTATATTCGCTCGCGGATCACAGTACCA 423
QY 139 AspGluLeuGluProLeuAlaIlePserAspAspGlyValValMetGlyLeuArgHisArg 158
DB 424 CCCAGCATGGAAGTCACATCTTCTAATCCGAGGGAGTTTGATGTGATGATCATCGC 483
QY 159 GluAspProLeuThrGlyValGlnPheHisProGlySerIleGlySerAspPheGlyArg 178
DB 484 AGCAGCCCATGTGGAGTGCAGTTCATCCGGAATCGATTGCGCGCATTCGGTGTG 543
QY 179 GluIleMetAlaAsnPheAspLeuAlaLeuAlaHisIleValArgAlaArgHisGly 198
DB 544 GAGATCATTTGACCGCTTTGTTGACCTATGCACACCGCATGATCGC-----588
QY 199 AlaAspSerProTyrgluLeuHisValArgArgValAspAlaLeuProAspAlaGlu 218
DB 589 -----ACCGATGAA 597
QY 219 ValArgArgGlyCysLeuProGlyGlyGlyThrThrPheThrLeuAspSerSerVal 238
DB 598 GTGAGCTTGTGCTGCAGCCCTGTGAG-----624
QY 239 LeuGluGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGlyTy 258
DB 625 -----CTTTCCATGCGTTGGC---GCGAGGAGACCATGCTGAATT 666
QY 259 LeuThrTyArgValAlaAspGlyValIleSerValArgGlySerAspGlyThrThr 278
DB 667 -----GAAAGCAGCGCTATGATTCGATTCGATTCGAGTGGCAGGTACCG 708
QY 279 ArgThrArgArgProPhePheAsnTyrlleuGlnGlnLeuGluArgArgValPro 298
DB 709 -----CACACATTTGAGAG---CTGAAAGTTTCTGATTTTCG 744
QY 299 ValAlaProGluLeu-----ProPheGluPheAsnLeuGlyTyValGlyTyLeu 315
DB 745 GTGAGCCCGGAGCGATGCGCGCGCG-----GTTGTGATCGGTACATT 789
QY 316 GlyTyGlyLeuLeuValaGlnThrThrGlyAspProAlaHisArg-----SerPro 332
DB 790 GGGTACGAGGCTAACGACGCTTCTTGGCATGCGGTCCATCGGCCAAGCCGCGGAA 849
QY 333 HisProAspAlaIlePheLeuPheAlaAspArgAlaIleLeuAspHisGlnGly 352
DB 850 TTTCACACGACTCCATGATGTATGACGCGAAATTATTCGATC-----CGTGGG 900
QY 353 CysCysTyrlleuLeuAlaLeuAspArgArgGlyHisAspAspGlyValAlaArgAlaTrp 371
DB 901 -----GATCGTCCCAATACCGCGCGCGGATTCGCGCTGGCGG 939
QY 372 ---LeuArgGlyThrAlaGlnThrLeuThrGlyLeuAlaValArgAlaProAlaGluPro 390
DB 940 CGGTACCGGAT-----GCGGTCTGCGCGCGGTCAAAAGC 975
QY 391 ThrProAlaMetValPheGlyIleProGluAlaAlaAlaGlyPheGlyProLeuAlaArg 410
DB 976 GTGCCACCGGTGCGAGTTCAACCG-----ACTGGTATCGGCGGTG---CAC 1023
QY 411 AlaArgHisAspAspAlaTyrlleuTyLeuTyLeuAspGlyCysLeuTyGlnIleArg 430
DB 1024 GTACGGGATCCCGAAGACGCTACATGCTACGATGACGCGATCCAAAGACGATTCCA 1083
QY 431 AsnGlyIleuSerTyrlleuLeuCysLeuThrAsnMetValThrAlaProThrGluAlaThr 450

```



```
Db      1084 GCCGGGAAACATATGAACTCTGACCAACCACTATTCGCCGAAGTACACGGTGG 1143
      451 AlaleuPro-----LeuTyrSerAlaleuArgAlaleuSerProValProTyrGlyAla 468
      1144 GTCCATCCGGCGGCAATGATCAAGCGCTAGACACCGCACTACTGCGCTATGCGCTGG 1203
      469 LeuLeuGluPheProGluLeuSerValLeuSerAlaSerProGluArgPheLeuThrIle 488
      1204 CTGGTTGTACCGATGATGATGACAGTGAATTTCCGCGTCAACCAAGACATTTATCACCATG 1263
      489 GlyAlaAspGlyValGluSerLysProIleLysGlyThrArgProArgGlyGlyThr 508
      1264 AAC---GACCCGATGTGTGTCTCTCCATCCATCAAGAAACCAAGAAACCCACGCGCAG 1320
      509 AlagluGluAspGluArgLeuArgAlaAspLeuAlaglyArgLysAspArgAlaglu 528
      1321 CGGAAAGAAAGACCGCGCTTGCGCGACGATCTGGACCAACCCCAAGACCGCGCGGAA 1380
      529 AsnLeuMetIleValAspLeuValArgAsnAspLeuAsnSerValCysAlaIleGlySer 548
      1381 AACCTATGATCGTGCACCTTGTTGGCGAAGACCTGCGCGCTGCGAATCAGCGAGC 1440
      549 ValHisValProArgLeuPheGluValGluThrTyrAlaProValHisGluLeuValSer 568
      1441 GTCCGTGTGCGCAAGATATGTGCACGACGCTTTCACACCGTGCACGACGATTTCT 1500
      569 ThrIleArgGlyArgLeuArgProGlyThrSerThrAlaAlaCysValArgAlaAlaPhe 588
      1501 ACCGTGAAAGGGCAACTTCGCGCCACCGACATGCTATGACGCTCCACGCGCACCTTC 1560
      589 ProGlyGlySerMetThrGlyAlaProLysLysArgThrmGluIleIleAspArgLeu 608
      1561 CCAAGCGGCTTCATGATGGGGGCGCCCAACCGCACCTGACCTCATCACCAGACTC 1620
      609 GluGluGlyProArgGlyValTyrSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAla 628
      1621 GAAGGCAAAACAACGCGGGTTTACTCCGATGATGATGATATGCGGACGATCTCCGC 1680
      629 AlaAspLeuSerIleValIleArgThrIleValLeuAlaAspGlyGlnAlaGluPheGly 648
      1681 ACCGATCTTCATGATGATCCGACCGCTGTCTCACACCCACCACTTGAGCTATGGG 1740
      649 ValGlyValIleValLeuSerLeuSerAspGluGluGluPheThrGluThrValVal 668
      1741 GTAGGAGGTCCATCATCGGCGCTTTTCAGACCCCGCAAGATGGGAGAAATCACACACA 1800
      669 LysAlaArgAlaMetValThrAlaLeu 677
      1801 AAATCCCGGTGCTTCTCGACCTACTC 1827

Db      1801 AAATCCCGGTGCTTCTCGACCTACTC 1827

RESULT 14
US-10-282-122A-16977
; Sequence 16977, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EUI/TA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; PRIORITY FILING DATE: 2003-02-20
; PRIORITY APPLICATION NUMBER: 60/191,078
```

```

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16977
; LENGTH: 1906
; TYPE: DNA
; ORGANISM: Clostridium difficile
US-10-282-122A-16977

Alignment Scores:
Pred. No.: 1,51e-89 Length: 1906
Score: 1001.00 Matches: 253
Percent Similarity: 52.52% Conservative: 112
Best Local Similarity: 36.40% Mismatches: 245
Query Match: 28,02% Indels: 85
DB: 17 Gaps: 20

US-10-089-514-2 (1-686) x US-10-282-122A-16977 (1-1906)
      4 LeuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGluThrTyrIleGlyAla 23
      4 CTATATGACATATGATATATATCTTTGTATACACCTGATCCAAATATATA--GAAAGAA 60
      24 ThrGlyGluProProValValValProAsnAspAlaAspTyrSerArgLeuProValGlu 43
      61 TTAGGGGAAACAGTTTATGATAAAGAAAT-----AATGAATCAAAATATAGT 108
      44 AspPhe-----AspAlaIleValSerProGlyProGlySerProAsp 58
      109 GACATTGAAGAACTAAATCCAGAGGTATGTATGTTATCTCCAGACCGGTCTCCA--- 165
      59 ArgGluArgAspPheGlyIleSerArgArgAlaIleThr-----AspSerGlyLeuPro 76
      166 -----AAAGAAAGCTGTGATTTGCAATGATATTTGTGACATTTTAAGGGAAAAACCT 219
      77 ValLeuGlyValCysLeuGlyHisGluGlyIleAlaGlnLeuPheGlyGlyThrValGly 96
      220 ATACTAGCAATTTGTTAGTGCACCAACTATTTGACAGCTTTTGGAGGAGATATTATA 279
      97 LeuAlaProGluProMetHisGlyArgValSerGluValArgHisThrGlyGluAspVal 116
      280 AAAGCACACAGCGCTGTGATGAGAAAGATATAGTATCAACCATCTAATTAAGAGATT 339
      117 PheArgGlyLeuProSerProPheThrAlaValArgTyrHisSerLeuAla-----Ala 134
      340 TTAGGGGACCTTAATAATCTTTAAATGTCACCTAGATATCATTCATAATTAATTGATTCC 399
      135 ThrAspLeuProAspGluLeuGluProLeuAlaTyrSerAspAspGlyValValMetGly 154
      400 AATACAGTTCCAAAAGAAATTAAGAAATTAACAGCAATTCATGAATAAGGAAATATGGGG 459
      155 LeuArgHisArgGlyLysProLeuThrGlyValGlnPheHisProGluSerIleGlySer 174
      460 ATTAGCATTAAGAAATATTAATTAAGAAAGATACAGTTTCATCCAGAACCTATTTGTGCA 519
```







QY 24 ThnGlyInProValValProAsnAspAlaAspTrpSerArgLeuProValGlu 43  
DB 375587 AAGGAAGATGTCGGGAGTGTCTGTAGAGAT-----TTTCGCTTCAA 375543  
QY 44 AspPhaAsp-----AlaIleValValSerProGlyProGlySerProAsp 58  
DB 375542 GCATTCACCAATTAGCAGCTGAAGATGATTAATCTTCGCCGACCTGGATTCGCGAT 375483  
QY 59 ArgGluArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeu 78  
DB 375482 -----GATTTCCAGAAAGTTAGCTCTCTCTAGACAAATTACGCTCCCATTTCTC 375432  
QY 79 GlyValCysLeuGlyIleGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAla 98  
DB 375431 GGGATTTGCTTCGCGCCATCAAAATGATGTCATTTTTCGGGCAAAAGTTGTTACAGCGC 375372  
QY 99 ProGluProMetHisGlyArgValSerGluValArgHisThrGlyGlyAspValPheArg 118  
DB 375371 GACGTTCTGTTTCAGGAAAGTATGATATCTCGCATCAGGTGAAGATTATTTGCC 375312  
QY 119 GlyLeuProSerProPheThrAlaValArgTyHisSerLeuAla-----AlaThrAsp 136  
DB 375311 GCACTTGATTCGAGATTTCTAGTTCGCCGCTTACCATTTCTTAGTCATTGACCCCAATACT 375252  
QY 137 LeuProAspGlyLeuGluProLeuAlaTrpSerAspAspGlyValValMetGlyLeuArg 156  
DB 375251 GTCCACAGCTGAATTGAAGTACGCGCTGTAAACAGATGATGGGCTTTTAATGGGGTTAGAG 375192  
QY 157 HisArgGlyLeuProLeuThrGlyValGlnPheHisProGlySerIleGlySerAspPhe 176  
DB 375191 CATCTTAACAAAGGATTTACAGTGTCAATTCATCTTAACCAATCTTCTTGAAAC 375132  
QY 177 GlyArgGlyIleMetAlaAsnPheArgAspLeuAlaLeuAlaHisHisArgAlaArgArg 196  
DB 375131 GGAACAGCATTTCTAGAAACTT----- 375108  
QY 197 HisGlyAlaAspSerProTyGlyLeuHisValArgArgValAspValLeuProAspAla 216  
DB 375108 ----- 375108  
QY 217 GluGluValArgArgGlyCys-LeuProGlyGluGlyThrThr-PheTrpLeuAspSers 236  
DB 375107 -----ATACGACTGGGGAGAGCTGCCAAATGAGTTTATTACGTTTGAATTTGAAGGTT 375054  
QY 236 er---SerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProL 255  
DB 375053 CGGTAAAGATTTTGAAGAACCAATTCGCACTGGGCGACCGCAT-----T 375006  
QY 255 euAlaGlyTyLeuThrTyArgValAlaAspGlyValValSerValArgGlySerAspG 275  
DB 375005 TATCGGAATGACTT----- 374992  
QY 275 LyrThrThrThrArgThrArgArgProPhePheAsnTyLeuGluGluGlnLeuGluArgA 295  
DB 374991 -----CTATAATGAAGCGCGCGAAGAACCCCAAAATTCAG 374955  
QY 295 rGArgValProValAlaProGlyLeuProPheGlnPheAsnLeuGlyTyVal---GlyT 314  
DB 374954 GAAA-----TACGTGCTGCTT 374937  
QY 314 TyrLeuGlyTyGlyLeuLysAlaGlnThrThrGlyAspProAlaHisArgSerProHisP 334  
DB 374936 TTGTGTATTATGAG-----GCCGCTCTGCTTTTCGTAAC----- 374902  
QY 334 roAspAlaAlaPheLeuPheAlaAspArgAlaIleAlaLeuAspHisGlnGluGlyCysC 354  
DB 374902 ----- 374902  
QY 354 ystTyLeuLeuAlaLeuAspArgArgGlyHisAspAspGlyAlaArgAlaTrpLeuArgG 374  
DB 374902 ----- 374902  
QY 374 luthrAlaGluThrLeuThrGlyLeuAlaValArgAlaProAlaGluProThrProAlaM 394

DB 374901 -----AATTTAAAAACCAACACCTGCGGAGAACATGCCACTTA 374862  
QY 394 erValPheGlyLe-----ProGluAlaAlaGlyPheGlyP 407  
DB 374861 TCTGGTTTGGTATGATCGATTAACCTTCAGACTGCACCTGCACGAAACCTAGACACTGCAC 374802  
QY 407 roLeuAla--ArgAlaArgHisAspLysAspAlaTyLeuLysArgIleAspGluCysL 426  
DB 374801 CACTTCTTTTAAATGATGATGAGGATTTACCCGAAATTAATCGGAAAAAATAGCAGACATTA 374742  
QY 426 euLysGluIleArgAsnGlyGlySerTyGlyIleCysLeuThrAsnMetValThrAla- 445  
DB 374741 AAGCAGAAATTCCTACTCGGAAACACATACCAATCAACTACCTGTTCCGCTCCAAAGTG 374682  
QY 446 -----ProThrGluAlaThrAlaLeuProLeuTySerAlaLeuArgAlaIleSerPro 464  
DB 374681 AAGTACCAAGTGAATTTCTGCCCAAGCTTCTTTGAATCGCTGCACAAAGTTGGAAG 374622  
QY 464 aLProTyGlyAlaLeuLeuGluPheProGlyLeuSerValLeuSerAlaSerProGluA 484  
DB 374621 CAACCTATACAGCGCTACTTGAAACGAGATTTTCAGATTAATCTCCGCTTACCCGAGAC 374562  
QY 484 rGPheLeuThrIleGlyAlaAspGlyGlyValGlnSerLysProIleLysGlyTyThrArgP 504  
DB 374561 TATTTTAAAA--TGGAAAGAAACATCTTACACGACAGTCCAAATGAAAGCACCATTC 374505  
QY 504 roArgGlyGlyThrAlaGluGluAspGluArgLeuArgAlaAspLeuAlaGlyArgGluL 524  
DB 374504 GCCGCGAAAGTACGGAACAAGCTGACTTGAAAGCACATGATTTGTTAAAAATGATTTCTA 374445  
QY 524 ysaAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAspLeuAsnSerValC 544  
DB 374444 AAATCGTCTGTAATAATGATGATTTGTTGATTTACTAGAAATAGACCTTGGGGTGATGT 374385  
QY 544 ysaAlaIleGlySerValHisValProArgLeuPheGlyValGluThrTyAlaProValH 564  
DB 374384 CGGTGCTGTAGCGTAAGAGTCCTAATTAATGACTTTAGAACCTTACCCACAGTGT 374325  
QY 564 leGlnLeuValSerThrIleArgGlyArgLeuArgProGlyLysSerThrAlaAlaCysV 584  
DB 374324 GGCAAATGACGCTTACTGTACCGGAGAAACACCTCTGATCTAGTTTAACCGCAGTTT 374265  
QY 584 aLArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLysLysArgThrMetGluI 604  
DB 374264 TCGAGCGCTTTTCCCTTGTGTTCAATACTGGCGCCCAAAAGCCGAGACATGGA 374205  
QY 604 leIleAspArgLeuGluGluGlyProArgGlyValTySerGlyValaLeuGlyTyTrpPhe 624  
DB 374204 TTATTTCAAGACTAGAAAGATTCACCTCGAGCGGTATACGTGGCGGATTTGATTCTTAG 374165  
QY 624 laLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleValLeuAlaAspGlyG 644  
DB 374144 AACCTAATGAAATGCGATTTTAAATGTCCAATTCGAACGATTTCCATTCAGATTAATA 374085  
QY 644 lnaAlaGluPheGlyValGlyGlyAlaIleValSerLeuSerAspGlnGluGluPheT 664  
DB 374084 AAGCAACTATAGTGTGTGTCGCGATATCTTTGGATTTCTGATGCCGAGACGAATTTT 374025  
QY 664 hArgThrValValLysAlaArgAlaMetValThrAlaLeuAspGlySerAlaVal 682  
DB 374024 CCGAGATTCATGCAAAATACAGCCATTTTGAAGAAACCAAGTTTCTTTAAAT 373969

Search completed: October 6, 2005, 00:28:12  
Job time : 20108.7 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 09:03:04 ; Search time 7504.63 Seconds  
(without alignments)  
3479.464 Million cell updates/sec

Title: US-10-089-514-2  
Perfect score: 3573  
Sequence: 1 MRTLIDNYSFTNHLFOYI.....VVKRAMVTALDGSAAVAGAR 686

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 34239544 seqs, 19032134700 residues  
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=cg2\_1/USFTO\_spool\_h/US10089514/runat\_04102005\_105744\_8043/app\_query.fasta\_1.1621  
-DB=EST -QFMT=fastcap -SUFFIX=ret -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pcc -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=US10089514 @CCN 1.1 6050 @runat\_04102005\_105744\_8043 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -MAIN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hic.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_gss1.\*  
9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	585	16.4	698	5	BQ625203
2	576	16.1	894	7	CO024641
3	566	15.9	859	2	BE658136
4	566	15.8	768	7	CO980064
5	562	15.7	818	7	CO030342
6	561	15.7	820	7	CO034510
7	556.5	15.6	893	9	CG942457
8	553	15.5	529	6	CA003778
9	547.5	15.3	620	2	AW223881

10	541.5	15.2	645	7	CV099881	CV099881	PANU_USDA
11	535.5	15.0	750	7	CO028609	CO028609	ESTR06593
12	525.5	14.7	579	7	CF598037	CF598037	NCEST3e06
13	511	14.3	528	2	AM509018	AM509018	e139b01.y
14	500	14.0	890	7	CF703510	CF703510	CCAFB0970
15	500	14.0	906	7	CF704316	CF704316	CCACN3070
16	486	13.6	1007	9	CNS065L5	CNS065L5	AL413391
17	483	13.5	523	8	B2412600	B2412600	OGAAG11TM
18	483	13.5	611	8	B2412593	B2412593	OGAAG11TC
19	483	13.5	631	8	B2537360	B2537360	OGAGD95TC
20	483	13.5	737	9	CG971361	CG971361	MBEUK12TF
21	483	13.5	810	9	CG170785	CG170785	PUIIR58TD
22	473	13.2	658	7	CV128706	CV128706	MdfrC3089
23	473	13.2	716	6	CD407015	CD407015	CM CK3199
24	472	13.2	540	4	BI427109	BI427109	saF76e01
25	468.5	13.1	773	7	CF718138	CF718138	CCAHB50TF
26	450.5	12.6	1954	3	CNS0A8YM	CNS0A8YM	EX818469
27	450.5	12.6	2015	3	CNS0A8YD	CNS0A8YD	Arabidops
28	446.5	12.5	660	8	B2892725	B2892725	Hm7_0152
29	443	12.4	1521	3	AY105009	AY105009	Zea_mays
30	437.5	12.2	653	7	CF863831	CF863831	PERZS008XG
31	436	12.2	845	8	B2570522	B2570522	mbH2_1411
32	433.5	12.1	502	4	BG508630	BG508630	bae75d07
33	419	11.7	2450	3	AY104124	AY104124	Zea_mays
34	417	11.7	657	9	CG947605	CG947605	MBEGN88TF
35	415	11.6	643	2	BE823672	BE823672	PAC81-60
36	414	11.6	1259	8	B2557654	B2557654	Arabidops
37	399	11.2	1873	3	CNS0A1AN	CNS0A1AN	Arabidops
38	392.5	11.0	1043	8	B2548859	B2548859	pac81-60
39	389	10.9	655	9	CG919310	CG919310	MBRAL61TR
40	388.5	10.9	770	7	CF703416	CF703416	CCAGD73TF
41	386	10.8	766	7	CF678009	CF678009	CCAHM30TF
42	385	10.8	429	9	AG258434	AG258434	locus_cor
43	384.5	10.8	668	6	CA095675	CA095675	SCCCL600
44	384.5	10.8	825	8	CC144065	CC144065	NDL_75A11
45	383.5	10.7	679	7	CN926353	CN926353	000430AEP

#### ALIGNMENTS

RESULT 1  
LOCUS BQ625203 698 bp mRNA linear EST 01-JUN-2002  
DEFINITION USDA-FP\_02294 Ridge pineapple sweet orange seedling Citrus  
ACCESSION BQ625203  
VERSION BQ625203.1 GI:21652372  
KEYWORDS EST.  
SOURCE Citrus sinensis  
ORGANISM Citrus sinensis  
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Sapindales; Rutaceae; Citrus.  
1 (bases 1 to 698)

#### REFERENCE

Bausher, M., McKendree, W., Dang, P., Chaparro, J., Shatters, R.,  
Hunter, W., and Niedz, R.  
Expressed sequence tags isolated from entire sweet orange (C.  
sinensis L. Osbeck) seedling  
Unpublished (2003)

#### JOURNAL

Contact: Michael Bausher  
US Horticultural Research  
USDA - ARS  
2001 South Rock Rd., Fort Pierce, FL 34945, USA  
Tel: (772) 462-5918  
Fax: (772) 462-5961  
Email: mbausher@uhrl.ars.usda.gov  
Seq primer: T3 Primer.

#### FEATURES

source location/Qualifiers  
1..698  
/organism="Citrus sinensis"  
/mol\_type="mRNA"  
/cultivar="Ridge pineapple"  
/db\_xref="taxon:2711"







FEATURES	source
Db	564 HieginleuVal1SerThrIleArgL1yrgLeuAArgProGly1YThSerThrAlaIleAcys 583
Db	373 CATCAAGCTCTGCACACCACTTCAATCCACATCGCCCCAACGTCGGCGGCTCCAGGTT 314
Qy	584 ValAArgAlaAlaPheProGly1YSerMetThrGlyAlaProLysArgThrMetGlu 603
Db	313 CTTCAGCGATGCTTCCACACGAGATTCATGACGGCGCTCCCAAACTCAGAGCGGTGCA 254
Qy	604 IleIleAaPaArgLeuGluGlu---GlyProAArgGlyVal1YrSerGlyAlaLeuGlyTrp 622
Db	253 ATCTTCGACGGCTCTGGAAAGAGACCCGACGCGCGCATATCTCCGACGATCTGGGCTAT 194
Qy	623 PheAlaIleuSerGlyAlaAlaAspLeuSerIleVal1IleArgThrIleValIleuAlaAsp 642
Db	193 GTATGCCCGCAGGAGACCGCTCGATATGTCGGGGTTATCCGACATTGTAAAGTACGGA 134
Qy	643 GlyAlaIleAgluphneGlyVal1Gly1YAlaIleVal1SerLeuSerAspGlnGluGlu 662
Db	133 AAGCAGCTTGAGCTAGTGTCGTGGAGGGCGGANTTCTGGCTCAGGAGCGGAGAAAGA 74
Qy	663 PheThrGluThrVal1Val1IleAlaArgAlaMetVal1ThrAlaLeu 677
Db	73 TGGCATGAGCTCATGCTGAGGCGCAATCTGTCTACTCGGCGCTG 29
RESULT 3	
BE658136/c	
LOCUS	BE658136 859 bp mRNA linear EST 24-MAY-2001
DEFINITION	GM700005A10A7 Gm-r1070 Glycine max cDNA clone Gm-r1070-1549 3',
ACCESSION	BE658136
VERSION	BE658136.1 GI:9984028
KEYWORDS	EST.
SOURCE	Glycine max (soybean)
ORGANISM	Glycine max
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosoid 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE	1 (bases 1 to 859)
AUTHORS	Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
TITLE	Expelling,J., Ragh.C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
JOURNAL	A Functional Genomics Program for Soybean (NSF 9872565)
COMMENT	Unpublished (1999)
	Other ESTs: A1940862 corresponding to Gm-cl010-1280 (5') Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565) Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics University of Illinois Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA Tel: (217) 244-6147 Fax: (217) 333-4582 Email: l-vodkin@uiuc.edu This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genome systems.com web site: www.genomesystems.com Seq primer: 5'-TTTCTTTTCTTTTCTTTTCTTTT(A/C/G)-3'.
	Location/Qualifiers 1..859

```

/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/cclone="Gm-r1070-1549"
/cclone_1b="Gm-r1070"
/note="The library Gm-r1070 is a sequence-driven, rereaked set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 869 from young pods. The 5' BstI of the source clones from the different libraries was used to select c1ones, or

```

a representative of each contig, which were rereacted to form library Gm-r1070. The cDNA clones of the rereacted Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the Laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, <http://www.cbc.umn.edu/Research/Projects/Soybean/Index.html>. Rereacting was performed by Genome Systems, St. Louis, <http://www.genomesystems.com>, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, <http://www.life.uiuc.edu/biotech/keck.html>. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

Alignment Scores:			
Pred. No.:	4.42e-43	Length:	859
Score:	567.00	Matches:	113
Percent Similarity:	72.55%	Conservative:	35
Best Local Similarity:	55.39%	Mismatches:	56
Query Match:	15.87%	Indels:	0
DB:	2	Gaps:	0
US-10-089-514-2 (1-686) x BE658136 (1-859)			
QY	474	GLULESERValleuSERAlaSERProGIuIRPhleuThRIleGIYAlaaspGIYGI	493
Db	797	GATCTGAGTANNNGCTNNNNNTNNNGAGAGTTCCTGCAGCTTGATAGCAACATATA	738
QY	494	ValGIuSERIySPRoIlleYSeGIYThRzPRProRGiYGIYThRIlaGIuGIuASpGIu	513
Db	737	CTAGAAGCTAAACCCATCAAGGAAAGCATCTCTGTCTACTGAAGAGAAAGTAG	678
QY	514	ArgIeuaRgAlaAspIeuaIaGIYArgIuIySaSPaRgaIaIuAenIuMeIleVal	533
Db	677	CAACTAAATTTAAATTACAGTTCAGCAAAAGATCAGGCTGAAAACCTGATATTGTT	618
QY	534	AspIeuaValaRgaSaSPleuSaSPeRValYsaIaIleGIySERValIHeValProArg	553
Db	617	GACCTTCTTAAGAAAGACCTTGCCCGCTGTGTGATCCTGGATCTGTTCATGTGCCAGT	558
QY	554	LeuPhEGIuValGIuThTRAlaProValIHeGIuEuaISeRThRIleArgIYArg	573
Db	557	CTCATGGATGTGAATCATATGCAACTGTGTACCACAATGTGATGATCTATTCTGGAGAG	498
QY	574	LeuArgPRoGIYThSeRThRIlaIaIaYsaValaRgaIaIaIHePRoGIYGIySeRMeC	593
Db	497	AAGCGTCAGATGTCAGTGTCTGTAGACGTGTCAAGCTGATTTCTCGGTGTTCAATG	438
QY	594	ThRIYAlaProIySeRyArgThMeGIuIleIleAsPaRgeuGIuGIuYProArg	613
Db	437	ACAGGTGCANNTAAGTTGAATGATCAATGSACTTCTTGATTTCTATTGAAAGTTGTTCCGA	378
QY	614	GIYValYySeRGIYAlaIeUGIYTRPhElaIeSeRGIYAlaIaIaAspIeuserIle	633
Db	377	GGATCTACTCAGCGGTGATTTGATTTTTCATATATATCAAGATTGATCTTAAATATT	318
QY	634	ValIleArgThRIleValIeuaIaSPGIYAlaIaGIuPhEGIYAlaGIYAlaIle	653
Db	317	GTGATTAAGAACAGTATTGTACACGAGGGTGAACTTCAATATAGAGCTGAGGGGCATTT	258
QY	654	ValSeRleuSeRAspGIuGIuGIuPhThRIgIuThRIYValYValYsaIaRgaIaMeC	673
Db	257	GTTCCTGTCTCAAAACCTGAAGACGAGATGAMAGATGTTTGAAGACAAAGCCCCA	198
QY	674	ValThRIaIeU 677	
Db	197	ACAAAGGCTGTG 186	



RESULT 4  
CO980064/c  
LOCUS  
DEFINITION  
CO980064  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

768 bp mRNA linear EST 13-SEP-2004  
Gm-r1089 Glycine max cDNA clone Gm-r1089-2405 3,  
mRNA sequence.  
CO980064  
GI:51336198  
EST.  
Glycine max (soybean)  
Glycine max  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 768)  
Vodkin, L., Shoemaker, R., Keim, P., Polacco, J. C., Retzel, E.,  
Khanna, A., Shealy, R., Clough, S., Thibaud-Nissen, F., Corbett, V.,  
Expediting, U., Gonzalez, D. O., Stromvik, M., Rodriguez-Huete, A. M.,  
Schweitzer, P., Gong, G. and Liu, L.  
A Functional Genomics Program for Soybean (NSF 9872565) (2004)  
Unpublished (2004)  
Other\_ESTRs: BE347129 corresponding to Gm-cl043-239 (5')  
Contact: Vodkin, L.O., PI, A Functional Genomics Program for  
Soybean (NSF 9872565)  
Lewin, H. A., Director, Keck Center for Comparative and Functional  
Genomics  
University of Illinois  
Ewin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA  
Tel: (217) 244-6147  
Fax: (217) 333-4582  
Email: l-vodkin@uiuc.edu  
Plate: Gm89007A1 row: C column: 03  
Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'  
High quality sequence stop: 768.  
Location/Qualifiers

FEATURES  
source

1. 768  
/organism="Glycine max"  
/mol\_type="mRNA"  
/db\_xref="taxon:3847"  
/clone="Gm-r1089-2405"  
/clone\_1db="Gm-r1089"  
/note="The library Gm-r1089 is a sequence-driven, reracked  
set of 9,216 low redundancy clones selected from 38  
different cDNA libraries constructed from various tissues  
and stages of development of soybean including 973 cDNAs  
from germinating cotyledons (source library Gm-cl069,  
Gm-cl076, and Gm-cl077); 1,465 cDNAs from various tissue  
and organ systems of the adult plant; 476 cDNAs from adult  
stem tissue (source library Gm-cl062); 1340 cDNAs from  
tissue culture derived somatic embryos (source libraries  
Gm-cl036 and Gm-cl075); 2918 cDNAs from hypocotyls or  
young seedlings; 742 cDNAs from germinating seedlings,  
shoot tips, or leaves exposed to various stresses (source  
libraries Gm-cl065, Gm-cl066, and Gm-cl068); 839 cDNAs  
from young leaves or hypocotyls exposed to bacterial and  
fungal pathogens (source libraries Gm-cl072, Gm-cl073,  
Gm-cl074; and Gm-cl084); and 463 from roots of young  
plants grown in hydroponic media without phosphate (source  
library Gm-cl087). The 5' ESTs of the source clones from  
the different libraries were used to select singletons, or  
a representative of each contig, which were reracked to  
form library Gm-r1089 and the cDNA clones of the reracked  
Gm-r1089 library were then sequenced at the 3' end. The  
unigene selection and 3' sequencing was funded by NSF  
Plant Genome project #9872565  
(http://soybeanomics.cropsci.uiuc.edu/) as part of  
creation of a low redundancy soybean cDNA set. The source  
cDNA libraries were constructed by the laboratories of  
Lila Vodkin, University of Illinois, Randy Shoemaker, Iowa  
State University, and Paul Keim, Northern Arizona  
University as part of the Public EST project,  
http://129.186.26.94/soybeanest.html. The contig analysis

to select unique genes was performed by the laboratory of  
Ernest Retzel, Center for Computational Genomics and  
Bioinformatics, University of Minnesota,  
http://web.abc.umn.edu/biodata/rsfoy/. Reracking and 3'  
sequencing were conducted by services of the University of  
Illinois Keck Center for Comparative and Functional  
Genomics http://www.biotech.uiuc.edu/keck.htm. Note: The  
corresponding 5' EST from each clone in the Gm-r1089  
library is listed in the 'OTHER EST' field. The detailed  
information on the source library for each clone can also  
be obtained by referring to the clone ID of the original  
cDNA library that is also listed under 'OTHER EST'."

ORIGIN

Alignment Scores:

Prod. No.:	4,66e-43	Length:	768
Score:	566.00	Matches:	113
Percent Similarity:	73.04%	Conservative:	36
Best Local Similarity:	55.39%	Mismatches:	55
Query Match:	15.84%	Indels:	0
DB:	7	Gaps:	0

US-10-089-514-2 (1-686) x CO980064 (1-768)

QY	474	GlUeUSeRValLeUSeRAlaSeRProGluARgPheUThrIleGlyAlaAspGlyGly	493
DB	745	GATCTGATATTGCTGCTCTNNNNNTGAGAGTTCCTCAGNNNGATGGAATATA	686
QY	494	ValGluSeRlyeProIleYseGlyThraRProaRgGlyGlyThraAlaGluGluAspGlu	513
DB	685	CTAGAAGCTTAAGCCCTCAAGGAAGATGATCGTGGCTTCACTGAAGAAGATGAG	626
QY	514	ArgLeuValAlaAspLeuAlaGlyArgGluYAspAspArgGluAsnLeuMetIleVal	533
DB	625	CAACTAAATTAAATTAACTGACGCGAAAGATCAGCGTGAATCTGATGTTGTT	566
QY	534	AspLeuValArgAsnAspLeuAsnSerValCysAlaIleGlySerValHisValProArg	553
DB	565	GACCTTCTAAGGATGATCCTTGCGCGTGTGTTGATCCGATGCTGATGCGACGT	506
QY	554	LeuPheGluValAlaGluThrTyraIleProValHisGluLeuValSerThrIleArgGlyArg	573
DB	505	CTCATGATGTGAATCATATGCACTGTTCAACATGCTGATGATCTGCTGGAAG	446
QY	574	LeuArgProGlyThraSerThrAlaIleCysValArgAlaIlePheProGlyGlySerMet	593
DB	445	AAGCGTCAAGATGCTGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG	386
QY	594	ThrGlyAlaProLysLysArgThrMetGluIleIleAspArgLeuGluGluProArg	613
DB	385	ACAGGTGACCTTAAGTGAATGATGCAATGCACTTCTGATTAAGAGTTGTTCTG	326
QY	614	GlyValTyrSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIle	633
DB	325	GGATCTACTCAAGCTGATGATGATTTCTCATTAATCAAGCACTTGTATTAATAT	266
QY	634	ValIleArgThrIleValLeuAlaAspGlyValIleGluPheGlyValGlyValIle	653
DB	265	GGATTAAGACAGCTATTTCAACGAGGTGAAGCTTCAATGAGAGCTGAGAGGCAATT	206
QY	654	ValSerLeuSerAspGluGluGluPheThrGluThrValValAlaAlaArgAlaMet	673
DB	205	GTTTGCTCTTCAACCTGAAGAAGGATGATGATGATGATGATGATGATGATGATG	146
QY	674	ValThrAlaLeu 677	
DB	145	ACAAGGCTGTGT 134	

RESULT 5  
CO030342/c  
LOCUS  
DEFINITION

818 bp mRNA linear EST 10-JUN-2004  
EST808726 Coccidioides posadasii spherule cDNA library, 0.5 to 5.3  
kb Coccidioides posadasii cDNA clone C1F368 3' end, mRNA sequence.



```

VERSION      CO030342
ACCESSION    CO030342.1   GI:48563318
KEYWORDS     EST
SOURCE       Coccidioides posadasii
ORGANISM     Coccidioides posadasii
              Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
              Onygenales; mitosporic Onygenales; Coccidioides.
REFERENCE    1 (bases 1 to 818)
AUTHORS      Gardner,M.J. and Cole,G.T.
TITLE        Analysis of gene expression in Coccidioides posadasii mycelia and
              spherules via expressed sequence tags
JOURNAL       Unpublished (2003)
COMMENT       Other ESTs: EST808727
               Contact: Gardner MJ
               The Institute for Genomic Research
               9712 Medical Center Drive, Rockville, MD 20850, USA
               Tel: 301 838 3519
               Fax: 301 838 0208
               Email: gardner@tigr.org.

FEATURES
             source
             1..818
                /organism="Coccidioides posadasii"
                /mol_type="mRNA"
                /strain="C735"
                /db_xref="taxon:199306"
                /clone="CIFB368"
                /dev_stage="spherules"
                /lab_host="E. coli DH10B, T1 phage resistant"
                /clone_lib="Coccidioides posadasii spherule cDNA library,
                0.5 to 5.3 kb"
                /note="Vector: pExpress 1; Site 1: Not I; Site 2: Eco RV;
                coccidioides posadasii spherule cDNA library, 0.5 to 5.3
                kb"

ORIGIN

Alignment Scores:
Pred. No.:          1,23e-42           Length:          818
Score:              562.00            Matches:         123
Percent Similarity: 62.98%            Conservative:    42
Best Local Similarity: 46.95%          Mismatches:     93
Query Match:        15.73%            Indels:         4
DB:                  Gaps:             3

US-10-089-514-2 (1-686) x CO030342 (1-818)

OY      420  LysAaPglIEaSPGlUCysLeuylSGlIIeaRgAsnGlyCUSeTyTgIInIEcyleu 439
Db      814  CAGAAAGATCCGGGCCCGCACAGAATGCCATTAAAGAGGCGCGATGAGTGACAATC 755
OY      440  ThraAmMeValThrLaPro---ThrgIuaIaThraLaeuProleuTyrsEraLaeu 458
Db      754  ACGACAAAATTCAGGCATCGTGTCCGGAGTAGACCACCATGCAATTATCTCTCCCT 695
OY      459  ArgAlaIleSerProValProtYrgIyaLaeuLeuGIuPhe-----ProGIueSer 476
Db      694  CGAGAACGTAACCCGCCGCCCTTACTCAGGGTATATAGACTTCGCTGCAACAGACACC 635
OY      477  ValLeuSerAlaSerProGIuaRgPheLeuThrIlIGlyAlaAPglYglYaGIuSer 496
Db      634  ATACTCTCTTCCTCGCCCGCAGAGGTTTATATCATGATGCTGATGGCGGCTGAATG 575
OY      497  LysProlIeLyGSglYThraRgProArGrGlyYThraIagluGIuaBPglUaRgLeuArg 516
Db      574  AAACCAATTAAGGGAACCTTGGCCGTCAGCCCCAGCAAAGAAAGACGAGCGAGAAAA 515
OY      517  AlaAPLeuAlIGlyARgGIuLySaPaRgaIaGIuaSnMeuElIeVaIaAPLeuVal 536
Db      514  AGTCAGTTAGCAGCGATGCAAGAGGCTCGCCGAGAACTTATGATGTCATCTTATC 455
OY      537  ArgAenAPLeuAnsSerValCyBaIaIlIGlySerValHisValProArGLEuPheGlu 556
Db      454  CGCGCGGACCTTCACAAACATTTCTCCATCAATAATCAAGGTCCGAAACTTATCCAC 395

```

OY		557	falGlnthrTyrAlaProValHisGlnIleuValSerThrIleArgGlyArgLeuArgPro	576
Dd		394	GTTGAACCTTCAGAAACATCATCAGCTCGTCAACACTTTCATCCCAATGCAGCCCC	335
OY		577	GlyThrSerThrAlaAcyValArgLaIalabPheProGlyGlySerMetThrGlyAla	596
Dd		334	AACGTGGGGGGCGGCCAGTTCTCTGAGCGAGTGCTTCCACGAGATTCATACGGGCGCT	275
OY		597	ProLysIlySarGthrMetGluIleIleAspArgLeuGluGlu---GlyProArgGlyAla	615
Dd		274	CCCAAACATCAGAGCGCGTCAATATCTCGACGCGCTCGAAGAAGCACCCGCAAGCGCGGCATA	215
OY		616	TyrsrGlyAlaLeuGlyTTPrpPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIle	635
Dd		214	TATCTCCGGCACTTCGGGCTATGATATGCCGCCAGCGCACCGTGCATCTCGGTGGATTATC	155
OY		636	ArgThrIleValIleuAlaAspGlyGlyAlaIagIubhegiValaGlyAlaIleValSer	655
Dd		154	CGTAGCATTTGAAAGTACGCAAGACGTTTAGCTAGGTGCTGGAGGGCGCATTTACTTGG	95
OY		656	IeuSerAspGlnGluGluIubPheThrGluThrValaIlySalAargAlaMerValThr	675
Dd		94	CTCAGCGAGCGGAGAAAGATGGAGATGAGTATCGTGAAGCGGAATGCTGTTCCTACG	35
OY		676	AlaIeu	677
Dd		34	GCCTG	29
RESULT 6				
LOCUS		C0034510	820 bp	mRNA linear EST 10-JUN-2004
DEFINITION		ESTR12894 Coccidioides posadasii spherule cDNA library, 0.5 to 5.3 kb Coccidioides posadasii CDNA clone ClFBU13 3' end, mRNA sequence.		
ACCESSION		C0034510		
VERSION		C0034510.1	GI:48571664	
KEYWORDS		EST.		
SOURCE		Coccidioides posadasii		
ORGANISM		Coccidioides posadasii		
		Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;		
		Omycelales; mitosporic Omycelales; Coccidioides.		
REFERENCE		1 (bases 1 to 820)		
AUTHORS		Gardner,M.J. and Cole,G.T.		
TITLE		Analysis of gene expression in Coccidioides posadasii mycelia and spherules via expressed sequence tags		
JOURNAL		Unpublished (2003)		
COMMENT		Other ESTs: ESTR12895		
		Contact: Gardner MJ		
		The Institute for Genomic Research		
		9712 Medical Center Drive, Rockville, MD 20850, USA		
		Tel: 301 838 3519		
		Fax: 301 838 0208		
		Email: gardner@igr.org		
FEATURES				
Source		Location/Qualifiers		
		1..820		
		/organism="Coccidioides posadasii"		
		/mol_type="mRNA"		
		/strain="C735"		
		/db_xref="taxon:199306"		
		/clone="ClFBU13"		
		/dev_stage="spherules"		
		/lab_host="E. coli DH10B, T1 phage resistant"		
		/clone_id="C0034510 Coccidioides posadasii spherule cDNA library, 0.5 to 5.3 kb"		
		/note="Vector: pExpress 1; Site_1: Not I; Site_2: Eco RV; Coccidioides posadasii spherule cDNA library, 0.5 to 5.3 kb"		
ORIGIN				
Alignment Scores:				
Pred. No.:		1,53e-42	Length:	820
Score:		561.00	Matches:	122
Percent Similarity:		64.03%	Conservative:	40
Best Local Similarity:		48..22%	Mismatches:	87



Query Match:	15.70%	Indels:	4
DB:	7	Gaps:	3
US-10-089-514-2 (1-686) x CO034510 (1-820)			



QY 534 AspleuVal1ArgAnaAspleuAsnSerValCysAlaIleGlySerValHisValProArg 553  
 Db 594 GACCTTCAAGAAATGACCTTGCCTGATGATCCTGATCTGTTGACGATCCGAT 653  
 QY 554 LeupheVal1ValGluThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArg 573  
 Db 654 CTCATGAGAGATCATATGATCAACCTTCCACCAATGGAGATCAATCTCGGGGAAA 713  
 QY 574 LeuArgProGlyThrSerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMet 593  
 Db 714 AAGAGTCAGATATTAAGTCTGTGACTGTGTCAAGTGCATTTCCGGGTTCATG 773  
 QY 594 ThrGlyAlaProGlySerGlyThrMetGluIleIleAspArgLeuGluGlyProArg 613  
 Db 774 ACAGGCCACCAAGTTCAGATCAATGGAATCTTGTCTCTCGAAAGTGTTCGA 833  
 QY 614 GlyValIleSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIle 633  
 Db 834 GGCATCTACTACAGCTCATTGGATT-TTCTCATATATCAACATTTGATCTAATATT 892

RESULT 8  
 CA003778 529 bp mRNA linear EST 23-OCT-2002  
 LOCUS HS15114r HS Hordeum vulgare subsp. vulgare cDNA clone HS15114

ACCESSION  
 CA003778  
 VERSION  
 CA003778.1 GI:24280760

KEYWORDS  
 EST.

SOURCE  
 Hordeum vulgare subsp. vulgare

ORGANISM  
 Hordeum vulgare subsp. vulgare  
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

REFERENCE  
 1 (bases 1 to 529)  
 Zhang,H., Potokina,E., Michalek,W., Weesche,W., Stein,N. and  
 Graner,A.

TITLE  
 JOURNAL  
 COMMENT  
 Barley ESTs from germinating seeds  
 Unpublished (2002)

CONTACT: Stein Nile  
 Molecular Markers Group, Department Genbank  
 Institute of Plant Genetics and Crop Plant Research (IPK)  
 Corrensstr. 3, 06466, Gatersleben, Germany  
 Tel: 039482-5522  
 Fax: 039482-5595

EMAIL: stein@ipk-gatersleben.de  
 Insert Length: 529 Std Error: 0.00  
 Plate: 15 row: 1 column: 14  
 Seq primer: M13rev.

FEATURES  
 source  
 Location/Qualifiers

1..529  
 /organism="Hordeum vulgare subsp. vulgare"  
 /mol\_type="mRNA"  
 /cultiivar="barke"  
 /sub\_species="vulgare"  
 /db\_xref="GABI:258990"  
 /db\_xref="taxon:112509"  
 /clone="HS15114"  
 /tissue\_type="embryo + scutellum"  
 /dev\_stage="0-16 hours after imbibition"  
 /lab\_host="XL10-Gold"  
 /clone\_lib="HS"  
 /note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA); Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable."

ORIGIN  
 Alignment Scores:  
 Pred. No.:

4.61e-42

Length:

529

Score: 553.00 Matches: 109  
 Percent Similarity: 76.00% Conservative: 24  
 Best Local Similarity: 62.29% Mismatches: 42  
 Query Match: 15.48% Indels: 0  
 DB: 6 Gaps: 0

US-10-089-514-2 (1-686) x CA003778 (1-529)

QY 496 SerIleProGlyIleGlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeu 515  
 Db 5 ACAGGCCAATCAAAAGGTATCAATATGACCTGGAGAACACCAAGAGAAAGTGTCTA 64  
 QY 516 ArgAlaAspleuAlaGlyArgGluValAspArgAlaGluAsnLeuMetIleValAspleu 535  
 Db 65 CGTTTGCACTGAAATACAGTAAAGACACCAAGCTGAGAACTTGATGATTTGATCTC 124  
 QY 536 ValArgAsnAspleuAsnSerValCysAlaIleGlySerValHisValProArgLeuPhe 555  
 Db 125 TTAAGAAATGATCTGGGTAAAGTCTCGACCTCGGAGCGTGCATGTTCTCGCTCATG 184  
 QY 556 GluValGluThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArg 575  
 Db 185 GAGGTGAAATCATATTAATCTGTTCAACCATGATGAGCACCATCCGGAGAAAGAAAG 244  
 QY 576 ProGlyThrSerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGly 595  
 Db 245 CTGATCTTAAGCCCGGTAGACTGCATCAAAAGCCCTTTCAGAGAGCTGATGACGGGT 304  
 QY 596 AlaProGlySerGlyThrMetGluIleIleAspArgLeuGluGlyProArgGlyVal 615  
 Db 305 GCCCGAAGGTCAAGTGCAGAGTGCAGATCTCGACCGCTTGAGAGACGCCAAGGGAATA 364  
 QY 616 TyrSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIle 635  
 Db 365 TACTCGGGGCGATCGGGTCTTCTGTCAACCGCACCGTTGCATGGAACATCGTATC 424  
 QY 636 ArgThrIleValLeuAlaAspGlyGlnAlaGluPheGlyValGlyAlaIleValSer 655  
 Db 425 AGACGGTATGCTGCACGACGGGTGCTCGCTCGTGCAGACGCGGGGATGTGCGC 484  
 QY 656 LeuSerAspGluGluGluGluPheThrGluThrValValIleValIle 670  
 Db 485 CTCCTAGACCCGAGCGGAGTACGCCGAGATATGCTCAAGCG 529

RESULT 9  
 AM223881 620 bp mRNA linear EST 18-MAY-2001  
 LOCUS EST100692 tomato fruit red ripe, TAMU Lycopersicon esculentum cDNA

DEFINITION  
 clone cLENI3P22, mRNA sequence.  
 AM223881

ACCESSION  
 AM223881  
 VERSION  
 AM223881.1 GI:6535565

KEYWORDS  
 EST.  
 SOURCE  
 Lycopersicon esculentum (tomato)

ORGANISM  
 Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicots; core eudicots;  
 asterids; lamids; Solanales; Solanales; Solanales; Solanum; Lycopersicon.  
 1 (bases 1 to 620)

REFERENCE  
 Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,  
 Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S.,  
 Roming,C.M., Frazer,C.M., Martin,G.B., Tanksley,S.D. and  
 Giovannoni,J.

TITLE  
 JOURNAL  
 COMMENT  
 Generation of ESTs from tomato fruit tissue  
 Unpublished (1999)  
 Contact: CUGI

LOCATION: University of California, Davis  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Email: <http://www.genome.clemson.edu/orders/index.html>  
 5 prime sequence.

FEATURES  
 source  
 Location/Qualifiers

1..620

/organism="Lycopersicon esculentum"



```

/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEN13P22"
/tissue_type="pericarp"
/dev_stage="red ripe (7-20 days post-breaker)"
/clone_lib="tomato fruit red ripe, TAMU"
/notes="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni; Fruit were tagged at the
breaker stage (first sign of lycopene accumulation on the
blossom end of the fruit) and harvested 7 days
post-breaker (fully red-ripe), 10 days post breaker, and
20 days post-breaker (over-ripe). 20 day fruit which
showed external or internal signs of pathogenesis were
discarded. Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."

```

## ORIGIN

## Alignment Scores:

Pred. No.:	1.94e-41	Length:	620
Score:	547.50	Matches:	112
Percent Similarity:	67.96%	Conservative:	28
Best Local Similarity:	54.37%	Mismatches:	63
Query Match:	15.32%	Indels:	3
DB:	2	Gaps:	2

US-10-089-514-2 (1-686) x AM223881 (1-620)

```

QY 419 LeuLYeArGILeAspGluCysLeuLYsGluILleArGAsnGlyGluSerTYrGluILeCys 438
   :::::::::::::::::::: ::::::::::::::::::::
DB 3 ATCAAAAGACGTTGAGAAATGTCTCAAGAGTTCTATAAAGAGAAAGTTATGAGTGTGT 62
   :::::::::::::::::::: ::::::::::::::::::::
QY 439 LeuThAsMeVAlThraLaProThrGluVala---ThraLaLeuProLeuTYrSerAla 457
   :::::::::::::::::::: ::::::::::::::::::::
DB 63 CTACAACTCAGATGAGAAATGAGAGTTGGGGGGAATAGATTCTCGAAGCTTATTCGTAAT 122
   :::::::::::::::::::: ::::::::::::::::::::
QY 458 LeuArGAlaILeSerProValProTYrGlyAlaLeuLeuGluDhe-----ProGluLeu 475
   :::::::::::::::::::: ::::::::::::::::::::
DB 123 CTCGAGATTAAGAAATCTCTGACCATATGCTGCTGCTTATTTTCAAGGAGAAACCTTA 182
   :::::::::::::::::::: ::::::::::::::::::::
QY 476 SerValLeuSerLaSerProGluArGPhaLeuThriLeGlyAlaSepGlyValGlu 495
   :::::::::::::::::::: ::::::::::::::::::::
DB 183 AGCATATGTTGTCATCACTGAAAGGTTCTTCATGATGACAGAAATGCTATTCTAGAA 242
   :::::::::::::::::::: ::::::::::::::::::::
QY 496 SerLYeProILeLYsGlyThraArGProArGlyGlyThraLaGluGluAspGluArGLeu 515
   :::::::::::::::::::: ::::::::::::::::::::
DB 243 GCAAAACCCATAAAGGACTATAGCTCGTGTTCCACCCCAAGAGAAATGATTTCTG 302
   :::::::::::::::::::: ::::::::::::::::::::
QY 516 ArGAlaAspLeuValaGlyArGluLYsAspArGAlaGluAsnLeuWecILeValaAspLeu 535
   :::::::::::::::::::: ::::::::::::::::::::
DB 303 AAACCTCAATTAGAAATGACAGTAAAGGATCAGCGGAAATTTGATGATTGTTGACTTG 362
   :::::::::::::::::::: ::::::::::::::::::::
QY 536 ValArGAsnAspLeuAsnSerValCysAlaILeGlySerValHisValProArGLeuPhe 555
   :::::::::::::::::::: ::::::::::::::::::::
DB 363 TTGAGAAATGACCTTGCGGCTGTATGTGAGACTGCGCTGTTCAATGCTCCACATCTCATG 422
   :::::::::::::::::::: ::::::::::::::::::::
QY 556 GluValaGILuThTYrAlaProValHisGluLeuValSerThriLeArGlyArGLeuArG 575
   :::::::::::::::::::: ::::::::::::::::::::
DB 423 GAATATGAATCCATATGACAGATTCCATGATGATGATGATGATGATGATGATGATGATGAT 482
   :::::::::::::::::::: ::::::::::::::::::::
QY 576 ProGlyThSerThraLaalaCysValaArGAlaAlaPheProGlyGlySerMetThrGly 595
   :::::::::::::::::::: ::::::::::::::::::::
DB 483 TCAGATGCAAGTCAATTGATGTGTGTTAGAGTGCATTCCTGCTGTGCTCAATGACAGGT 542
   :::::::::::::::::::: ::::::::::::::::::::
QY 596 AlaProLYeArGArGThrMetGluILeILeAspArGLeuGluGluGlyProArGlyVal 615
   :::::::::::::::::::: ::::::::::::::::::::
DB 543 GCACCAAAAGTTGAGATCAATGAATCTTGTATCATCTTGAAATTTGTTGAGGGGACATA 602
   :::::::::::::::::::: ::::::::::::::::::::
QY 616 TYrSerGlyAlaLeuGly 621
   :::::::::::::::::::: ::::::::::::::::::::
DB 603 TACTCGGGCTGCATTGGA 620
   :::::::::::::::::::: ::::::::::::::::::::

```

RESULT 10

CV099881  
LOCUS  
DEFINITION  
FAMU USDA FP 7904 Vitis shuttleworthii L., grape Vitis  
shuttleworthii cDNA clone WMS046\_D11 5', mRNA sequence.  
CV099881  
ACCESSION  
VERSION  
CV099881.1 GI:51582045  
KEYWORDS  
SOURCE  
ORGANISM  
Vitis shuttleworthii  
Vitis shuttleworthii  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosids; Vitaceae; Vitis.

## REFERENCE

1 (bases 1 to 645)  
Hunter, W.B., Dang, P.M., Chaparro, D.X., Lu, J., Huang, H. and Leong, S.  
Genes expressed in Vitis shuttleworthii L  
Unpublished (2004)  
COMMENT  
Contact: JIANG Lu, Wayne Hunter, Phat Dang and Hong Huang  
Genetic Lab, Viticulture Center, U.S. Horticultural Research Lab  
FAMU, USDA-ARS,  
Tallahassee, FL, Ft. Pierce, FL, USA  
Tel: (850) 412-7393, (772) 462-5898  
Fax: (850) 561-2617, (772) 462-5986  
Email: jiang.lu@fam.u.edu, hong.huang@fam.u.edu,  
whunter@ushrl.ars.usda.gov, pdang@ushrl.ars.usda.gov  
Seq primer: T3 Primer.

## FEATURES

## source

```

1..645
Location/Qualifiers
/organism="Vitis shuttleworthii"
/mol_type="mRNA"
/db_xref="taxon:246827"
/clone="WMS046_D11"
/sex="Mixed population"
/tissue_type="Entire tendril, leaves, bud, flowers"
/dev_stage="At blooming"
/lab_note="XLI-Blue"
/clone_lib="Vitis shuttleworthii L., grape"
/notes="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
XhoI; A high quality EST with at least 100 contiguous
bases at Trace Tuner score of 20 or better. Construction
by PM Dang, USDA, ARS, U.S. Horticultural Research Lab,
Ft. Pierce, FL, USA."

```

## ORIGIN

## Alignment Scores:

Pred. No.:	7.65e-41	Length:	645
Score:	541.50	Matches:	117
Percent Similarity:	70.56%	Conservative:	34
Best Local Similarity:	54.67%	Mismatches:	60
Query Match:	15.16%	Indels:	4
DB:	7	Gaps:	2

US-10-089-514-2 (1-686) x CV099881 (1-645)

```

QY 434 SerTYrGluILeCysLeuThraSMeVAlThraLaProThr--GluAlaThraLaLeu 452
   :::::::::::::::::::: ::::::::::::::::::::
DB 3 AGTTATGAATGTGTCTTCAACTCAATGAGAAAGAGATTGGCAATAGATTACTTG 62
   :::::::::::::::::::: ::::::::::::::::::::
QY 453 ProLeuTYrSerAlaLeuArGAlaILeSerProValProTYrGlyAlaLeuLeuGluPhe 472
   :::::::::::::::::::: ::::::::::::::::::::
DB 63 GGACTTATCTTAACCTTAGAGAAAAAATCCAGCACCATATGACACCTGCTTAATTTT 122
   :::::::::::::::::::: ::::::::::::::::::::
QY 473 ProGlu-----LeuSerValLeuSerAlaSerProGluArGPhaLeuThriLeGlyAla 490
   :::::::::::::::::::: ::::::::::::::::::::
DB 123 TCAAAACAAAACCTGTGCATCTGCTGCTTCTTACCAAGAGAGTTTCTACAGCTGATGCA 182
   :::::::::::::::::::: ::::::::::::::::::::
QY 491 AspGlyValaGluSerLYsProILeLYsGlyThraArGProArGlyGlyThraLaGlu 510
   :::::::::::::::::::: ::::::::::::::::::::
DB 183 AATGTAATTTTGAAGACCAAGCCCATCAAGGAGTATGCTGCTGTTGACGAAAGAG 242
   :::::::::::::::::::: ::::::::::::::::::::
QY 511 GluAspGluArGLeuArGAlaAspLeuValaGlyArGluLYsAspArGAlaGluAsnLeu 530
   :::::::::::::::::::: ::::::::::::::::::::
DB 243 GAAGATGAACACCTCAAAATTGCAATCAATACAGTGAAGAAAGATCAGGCTGAATCTG 302
   :::::::::::::::::::: ::::::::::::::::::::

```



Qy 531 MetIleValAspLeuValArgAsnAspLeuAsnSerValCysAlaIleGlySerValHis 550  
Db ATGATGTGTTGATCTTCAAGAAATGACCTTGCTGCTGATGAACTGGAATCTATTCAT 362  
Qy 551 ValProArgLeuPheGluValGluThrTyrAlaProValHisGlnLeuValSerThrIle 570  
Db GTTCTCTGCTTAAATGATGGAATGCAATGCAACTGTTCAATACCATGATGATGACAT 422  
Qy 571 ArgGlyLeuArgProGlyThrSerThrAlaIleCysValArgAlaAlaPheProGly 590  
Db CGAGAAAGAAAGCAATCCAGATGAGACCTGCAATGTCAGAGACGCTTCCAGGT 482  
Qy 591 GlySerMetThrGlyAlaProValArgThrMetGluIleLeuAspArgLeuGlu 610  
Db GGATCATGACAGTGGCCCGAAGTAAATGATGAACTTCTGATTCATGAAAT- 541  
Qy 611 GlyProArgGlyValTyrSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAsp 630  
Db AGTTCTAGAGGTATCTACTCTGCTGAGGCTTTCATCAACACAGACATTTGAT 601  
Qy 631 LeuSerIleValIleArgThrIleValLeuAlaAspGlyGln 644  
Db CTCAACATGTGATGAAGACGATGTCATGATGAGGTGA 643

RESULT 11  
LOCUS CO028609/c 750 bp mRNA linear EST 10-JUN-2004  
DEFINITION EST806999 Coccidioides posadasii spherule cDNA library, 0.5 to 5.3  
kb Coccidioides posadasii cDNA clone CIPAS82 3' end, mRNA sequence.  
VERSION CO028609  
KEYWORDS EST  
SOURCE Coccidioides posadasii  
ORGANISM Coccidioides posadasii  
REFERENCE Eukaryote; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Oomyzales; mitosporic Oomyzales; Coccidioides.  
1 (bases 1 to 750)  
AUTHORS Gardner M.J. and Cole G.T.  
TITLE Analysis of gene expression in Coccidioides posadasii mycelia and  
spherules via expressed sequence tags  
JOURNAL Unpublished (2003)  
COMMENT Other ESTs: EST806994  
Contact: Gardner MJ  
The Institute for Genomic Research  
9112 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301 838 3519  
Fax: 301 838 0208  
Email: gardner@igf.org

FEATURES  
source 1..750  
Location/Qualifiers  
1..750  
/organism="Coccidioides posadasii"  
/mol\_type="mRNA"  
/strain="C735"  
/db\_xref="taxon:199306"  
/clone="CIPAS82"  
/dev\_stage="spherules"  
/lab\_host="E. coli DH10B, T1 phage resistant"  
/clone\_lib="Coccidioides posadasii spherule cDNA library,  
0.5 to 5.3 kb"  
/note="Vector: pExpress 1; Site 1: Not I; Site 2: Eco RV;  
Coccidioides posadasii spherule cDNA library, 0.5 to 5.3  
kb"

## ORIGIN

Alignment Scores:  
Pied. No.: 3.55e-40 Length: 750  
Score: 535.50 Matches: 115  
Percent Similarity: 66.08% Conservative: 35  
Best Local Similarity: 50.66% Mismatches: 74  
Query Match: 14.99% Indels: 3  
DB: 7 Gaps: 2

US-10-089-514-2 (1-686) x CO028609 (1-750)

Qy 454 LeuTyrSerAlaLeuArgAlaIleSerProValProTyrGlyAlaLeuLeuGluPhe--- 472  
Db TTATATCTTCCCTTGAGAAAGTAACTGACCCCTTACTACAGGATATATGACTTCCT 650  
Qy 473 ---ProGluLeuSerValLeuSerAlaSerProGluArgPheLeuThrIleGlyAlaAsp 491  
Db GTCAAGAGAACACCATACTCTCTCTGCTGCCGAGAGGTTTATTCATCATGATGCTGAT 590  
Qy 492 GlyGlyValAlaLeuSerLysProIleGlyTyrArgProArgGlyGlyThrAlaGluGlu 511  
Db GGCGTGCTGAATAATGAACCAATAAAGGAACCTTGCCCTGACGCCAGCAAAAGAGAA 530  
Qy 512 AspGluArgLeuArgAlaAspLeuAlaGlyArgGlyLysAspArgAlaGluLeuMet 531  
Db GACGAGCGAGAGAAAGTCAAGTACGACGATGTCAGAGAGCTGCGCGAATCTTATG 470  
Qy 532 IleValAspLeuValArgAsnAspLeuAsnSerValCysAlaIleGlySerValHisVal 551  
Db ATTTGTCATCTTATCCGCGCGACCTCCACACATTTCTCATCAAAATCAATCAAGTTC 410  
Qy 552 ProArgLeuPheGluValGluThrTyrAlaProValHisGlnLeuValSerThrIleArg 571  
Db CCGAACTACTCCACGCTGAAAGCTAGAAACAGTACATCAGCTCGTCACACATTCMA 350  
Qy 572 GlyArgLeuArgProGlyThrSerThrAlaIleCysValArgAlaAlaPheProGlyGly 591  
Db TCCACATCCGCCCAACGTCGCGCGCTGAGTTCAGAGATGCTTCCACACAGGA 290  
Qy 592 SerMetThrGlyAlaProLysArgThrMetGluIleLeuAspArgLeuGlu--- 610  
Db TCCATAGCGGCGCTCCCAACTCAGAGCCGTGCAATCTTCACGCGCTGGAAGACAC 230  
Qy 611 GlyProArgGlyValTyrSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAsp 630  
Db CGGAGCGCGCATATATCTCCGCACTGAGCTATGTATGCCAGCGGACCGTCGAT 170  
Qy 631 LeuSerIleValIleArgThrIleValLeuAlaAspGlyGlnAlaGluPheGlyValGly 650  
Db CACTCGGTGATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 110  
Qy 651 GlyAlaIleValSerLeuSerLeuAspGluGluGluPheThrGluThrValValAla 670  
Db GGAGCGCATTTACTGCTGACGAGCGGAGAAAGATGAGTATGATGATGATGATGATGAT 50  
Qy 671 ArgAlaMetValThrAlaLeu 677  
Db AATGCTGTGCTACGCGCGCTG 29

RESULT 12  
LOCUS CF598037 579 bp mRNA linear EST 26-SEP-2003  
DEFINITION NCST3e06c12.v1 NC-LIV Tachyzoite cDNA library Neospora caninum  
cDNA 5' similar to TR:P72539 P72539 PAPA.; mRNA sequence.  
VERSION CF598037  
KEYWORDS EST  
SOURCE Neospora caninum  
ORGANISM Neospora caninum  
REFERENCE Eukaryote; Alveolata; Apicomplexa; Coccidia; Eimeriida;  
Sarcocystidae; Neospora.  
1 (bases 1 to 579)

AUTHORS Cole, R., Fogarty, S., Tang, K., Howe, D. K., Sibley, L. D., Clifton, S.,  
Matta, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Theising, B.,  
Bowers, Y., Gibbons, M., Ritter, E., Bennet, J., Ronko, I.,  
Teagarden, S., R., Pedele, M., Belaygorod, L., Franklin, C.,  
Carr, L. M., Grow, A., Maguire, L., Watkins, J., Richey, J., Waterson, R.  
and Wilson, R.  
TITLE USDA-WashU Neospora EST Project  
JOURNAL Unpublished (2000)  
COMMENT Contact: Sandy Clifton, Ph.D. - Neospora  
USDA-WashU Neospora EST Project  
Washington University School of Medicine







QY	4939	11lelysglyThrArgProarglylglYThrAlaaglunGuuaSpGluarGluenrgAlasp	518		
Db	1	ATCAAGGAAACGATGCTGCTGTGTCTACTGAGAGGAAAGATGACCAACTAAATT7AAA	60		
QY	519	LeuAlaGlyArgGluLysAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsn	538		
Db	61	TTACAGCTTCAGGAAAGAGATCAGGCTGAAAACCTGATGATGTTGACCTTCTAAGAAAT	120		
QY	539	AspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGlu	558		
Db	121	GACCTTGCCCGGTGTGTGATCTCTGATCTGTTCAATGTCACACCTCATGAGATGTGAA	180		
QY	559	ThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyThr	578		
Db	181	TCATATGCAACTGTTCACATGGGTGAGTACTATTCTGGGCAAAACGGTCAGATGTC	240		
QY	579	SerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLys	598		
Db	241	AGTGTGTAGACTGTGTCAAAGCTGCATTTCTGTGTGTTCAATACAGGTGCACCTAAG	300		
QY	599	LysArgThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTyrSerGly	618		
Db	301	TTGAGATCAATGGAACCTTGATTCATTCATGCAAAAGTTGTCGAGGATCATCAGGC	360		
QY	619	AlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIle	638		
Db	361	TGTATTTGATTTTCTCATATATATCAACCATTTGATTCGAATATTTGTATATAGAACACTG	420		
QY	639	ValLeuAlaAspGlyGlyAlaGluPheGlyValGlyGlyAlaIleValSerLeuSerasp	658		
Db	421	ATTGTACACGAGCGTGAAGCTTCATAGAGCTGACCGGCAGATGTTGCTGTCAAAAC	480		
QY	659	GlnGluGluGluPheThrGluThrValValIysAlaArgAla	672		
Db	481	ACTTGAGACGAGTATGAGAGATGGGTATGAAAGACAGCGG	522		
RESULT 14					
LOCUS	CF703510/c	890 bp	mRNA linear EST 16-AUG-2004		
DEFINITION	CAAFB09TO C.neoformans strain JEC21 Cryptococcus neoformans var.				
ACCESSION	CF703510	neoformans cDNA clone CCAF809, mRNA sequence.			
VERSION	CF703510.1	GI:41557669			
KEYWORDS	EST.				
SOURCE	Cryptococcus neoformans var. neoformans (Filobasidiella neoformans var. neoformans)				
ORGANISM	Cryptococcus neoformans var. neoformans				
REFERENCE	Eukaryota; Fungi; Basidiomycota; Hymenomycetes;				
AUTHORS	Heterobasidiomycetes; Tremellomycetidae; Tremellales; Tremellaceae;				
TITLE	Filobasidiella, 1 (bases 1 to 890)				
JOURNAL	Loftus, B.				
COMMENT	End sequencing of clones from a full length enriched, normalized JEC21 cDNA library				
	Unpublished (2003)				
	Other_Estis: CCAF809TR				
	Contact: Brendan Loftus				
	TIGR				
	Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,				
	Rockville, MD 20850, USA				
	Tel: 301-838-3543				
	Fax: 301-838-0208				
	Email: crypt@tigr.org				
	Seq primer: TF.				
FEATURES	Location/Qualifiers				
Source	1..890				
	/organism="Cryptococcus neoformans var. neoformans"				
	/mol_type="mRNA"				
	/strain="JEC21"				
	/db_xref="taxon:40410"				
	/clone="CCAF809"				
	/clone_jlb="C.neoformans strain JEC21"				
	/note="Vector: pCKVSPORT6; Site_1: NotI EcoRV; The full				

length, normalized library was prepared from a variety of conditions using RNA provided by Joseph Heltman and Jennifer Lodge"

Alignment Scores:	
Pred. No.:	1,076-36
Score:	500.00
Percent Similarity:	55.36%
Best Local Similarity:	41.79%
Query Match:	13.99%
DB:	7
Length:	890
Matches:	117
Conservative:	38
Mismatches:	83
Indels:	42
Gaps:	6

US-10-089-514-2 (1-686) x CF703510 (1-890)

[illegible]







GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 19:00:45 ; Search time 1433.22 Seconds  
(without alignments)  
10548.289 Million cell updates/sec

Title: US-10-089-514-3  
Perfect score: 312  
Sequence: 1 atgaccgagcagcagcagc.....cggcgcgagcgcgcgcgtga 312

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb da:\*  
2: gb hlg:\*  
3: gb\_in:\*  
4: gb\_com:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_dr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	312	100.0	312	6	BD178314 Transform
2	312	100.0	312	6	BD093915 Transform
3	312	100.0	5251	1	AB116234 Streptomy
4	296.4	95.0	14159	1	AF262220 Streptomy
5	135.6	43.5	387	6	A48325 Sequence 3
6	135.6	43.5	387	6	AR198355 Sequence
7	135.6	43.5	2888	6	A48323 Sequence 1
8	135.6	43.5	2888	6	AR198353 Sequence
9	135.6	43.5	4740	1	SPU60417 Streptomy
10	64	20.5	10029	1	AF263011S2 Streptomy
11	60.8	19.5	13635	1	AE005068 Halobacte
12	59.4	19.0	1964	8	AK121213 Cryza sat
13	58.2	18.7	16905	1	AE005050 Halobacte
14	55.8	17.9	110000	1	AP006840_33 Continuation (34 o
15	55.8	17.9	248550	1	SC0939120 Streptomy
16	55.4	17.8	300800	1	SC0939112 Streptomy
17	55.2	17.7	346259	1	BX640435 Bordetell
18	55	17.6	132544	1	AF521085 Streptomy
19	55	17.6	293900	1	AP005026 Streptomy

20	54.8	17.6	1458	6	AX573996 Sequence
21	54.8	17.6	45055	6	AX574197 Sequence
22	54.8	17.6	300800	1	AP005036 Streptomy
23	54.8	17.6	349028	1	BX640413 Bordetell
24	54.6	17.5	1349	4	BOVT2TRANS
25	54.6	17.5	73882	6	CQ870950 Sequence
26	54.6	17.5	89973	1	AJ605139 Actinopla
27	54.6	17.5	89976	1	AJ632270 Actinopla
28	54.6	17.5	110000	1	AP006618_28 Continuation (29 o
29	54.4	17.4	110000	1	AP006618_43 Continuation (44 o
30	54.4	17.4	17050	1	AL646081 Ralstonia
31	54.4	17.4	321250	1	SC0939111 Streptomy
32	54.4	17.4	343473	1	BX640451 Bordetell
33	54.4	17.4	346259	1	BX640435 Bordetell
34	54.4	17.4	349726	1	BX640421 Bordetell
35	54	17.3	1208	6	A37834 Sequence 4
36	54	17.3	1208	6	AR069872 Sequence
37	54	17.3	1208	6	AR099268 Sequence
38	54	17.3	1208	6	AR124153 Sequence
39	54	17.3	1208	6	AR442760 Sequence
40	54	17.3	5392	6	A37831 Sequence 1
41	54	17.3	5392	6	AR069869 Sequence
42	54	17.3	5392	6	AR099265 Sequence
43	54	17.3	5392	6	AR124150 Sequence
44	54	17.3	5392	6	AR442757 Sequence
45	54	17.3	190050	1	AL646080 Ralstonia

## ALIGNMENTS

RESULT 1  
BD178314  
LOCUS  
DEFINITION  
BD178314 Transformant producing pF1022 substance and process for producing the same and novel biosynthetic gene.  
ACCESSION  
BD178314  
VERSION  
BD178314.1 GI:30015579  
KEYWORDS  
WO 02077244-A/2.  
SOURCE  
Streptomyces venezuelae  
ORGANISM  
Streptomyces venezuelae  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
REFERENCE  
1 (bases 1 to 312)  
Yanai,K., Sumida,N., Watanabe,M., Moriya,T. and Murakami,T.  
Transformant producing pF1022 substance and process for producing the same and novel biosynthetic gene  
Patent: WO 02077244-A 2 03-OCT-2002;  
MORIJI SEIKA KAISHA LTD, KOJI YANAI, NAOMI SUMIDA, MANABU WATANABE, TATSUKI MORIYA, TAKESHI MURAKAMI  
COMMENT  
OS Streptomyces venezuelae  
PN WO 02077244-A/2  
PD 03-OCT-2002  
PR 22-MAR-2002 WO 2002JP002782  
PI KOJI YANAI, NAOMI SUMIDA, MANABU WATANABE, TATSUKI MORIYA, TAKESHI MURAKAMI  
PC C12N15/61, C12N1/15, C12P17/14// (C12N1/15, C12R1:645), (C12P17/14, C12R1:645)  
PC Transformant producing pF1022 substance and process for CC  
CC producing the same  
FH and novel biosynthetic gene  
FT Key location/Qualifiers (1). (309).  
CDS location/Qualifiers

## FEATURES

source  
1. 312  
/organism="Streptomyces venezuelae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:54571"

## ORIGIN

Query Match 100.0%; Score 312; DB 6; Length 312;



Best Local Similarity 100.0%; Pred. No. 4.1e-33;  
Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]

QY 61 CTCTGGACACGGTGGGGCCCGGCATGACCTCGGTGTCGCGATCGCGGTACAGTCC 12  
|||  
Dp 61 CTCTGGACACGGTGGGGCCCGGCATGACCTCGGTGTCGCGATCGCGGTACAGTCC 12

QY 121 CGGACGCGCCTCCCAATGATGACAGCCCGGCGTCAAGCTGGTCAAGGACAGGGCCGGCC 18

Dp 121 CGGACGCGCGTCCCGATGATGACAGCCCGGCGTCAAGCTGGTCAAGGACAGGGCCGGCC 18

Dy 181 CGCTACGCCCGCAGCACGAGCTGTCTTCTGGTGAACCTCTACGACGTATC 24  
Dc 181 CGCTACGCCCGCAGCACGAGCTGTCTTCTGGTGAACCTCTACGACGTATC 24

QY 241 ATCACGAGATGTGCGCGCTCAGAGACTGTGATAGCCGGAGAGCCTGACGGCCGAG 30  
|||  
DQ 241 ATCACGAGATGTGCGCGCTCAGAGACTGTGATAGCCGGAGAGCCTGACGGCCGAG 30  
|||

QY	301	GACCGCGCGTGA	312
Db	301	GACCGCGCGTGA	312

LOCUS	BD093915	312 bp	DNA	linear	PAT 27-AUG-20
RESULT 2	BD093915				

DEFINITION	transformants that produce secondary metabolites modified by a functional group(s) and novel biosynthesis genes.
ACCESSION	BD093915
VERSION	BD093915.1 GI:22639503

INDIVIDUALS	SOURCE	ORGANISM
NO 045342-74/2.		
	Streptomyces venezuelae	
	Streptomyces venezuelae	
	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	

REFERENCE  
1 (bases 1 to 312)  
Yanai, K., Okakura, K., Yasuda, S., Watanabe, M., Miyamoto, K., Mido, N. and Murakami, T.

functional group(s) and novel biosynthesis genes  
Patent: WO 0123542-A 2 05-APR-2001;  
MEIJI SEIKA KAISHA LTD, KOJI YANAI, KAORU OKAKURA, SHOHEI YASUDA,

COMMENT	OS	PN	PD
	Streptomyces venezuelae	WO 0123542-A/2	05-APR-2001

PR 29-SEP-1999 JP 99P 276314  
PI KOJI YANAI, KAORU OKAKURA, SHOHEI YASUDA, MANABU WATANABE, KOICHI  
PI MIYAMOTO,  
PI SHIGEKI UEDA, SHIGERU ICHIMARU

PC	C12N15/09, C12N5/10, C12P21/02, C07K11/00// (C12P21/02, C12R1:645
CC	
FH	
Key	Location/Qualifiers
PM	(1)

FEATURES	Location/Qualifiers
source	1. .312
	/organism="Streptomyces venezuelae"
	/mol_type="Genomic DNA"

Query Match	Score	DB	Length
100.0%	312.	DB 6.	Length 312.

1 ATGACCGAGAGAAACGAGCTGCGAGCGGCTGACGCGGAGCTTGGACGCGCTTGGACGGAAGC 60  
 Best Local Similarity 100.0%; Pred. No. 4.1e-33;  
 Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0

\_\_\_\_\_

D**b** 1 ATGACCGAGCAGAACGAGCTTCAGCGGCTGCCGCGGAGCTCGACGCCCTCGACGGGACG 60

Db 61 CTCTGGACACGGTGGCGCCGCGCATCGACTCGGTCGCGCATCGCGGGTACAAGTCC 12

121 CGGACGGCGTCCGATGATCAGCCCGCGGGTCAGCTGGTCAAGGACAGGGCGGCC 18

Db 181 CGCTACGCCCCGACCACGGCTCGACGAATGTTCTCGTGTAACCTTACGACGTGATC 24

Db 241 ATCAGGAGATGTGCCCGCTCGAGGACTGCTGATGACCCGGAGAGCCTGACGCCGAG 30

Db 301 GACCGCGGTGA 312

RESULT 3	AB116234	LOCUS	AB116234	5251 bp	DNA	linear	BCT 03-JUL-20
DEFINITION	Streptomyces venezuelae	nc_004424	nc_004424	5251 bp	DNA	linear	BCT 03-JUL-20

4-amino-4-deoxychorismate synthase, 4-amino-4-deoxyprephenate dehydrogenase, 4-amino-4-deoxychorismate mutase, hypothetical protein, complete cds.

ORGANISM	Streptomyces venezuelae
SOURCE	Streptomyces venezuelae
KEYWORDS	AB116234.1
VERSION	GI:47846865

REFERENCE

1  
Yanai, K., Sumida, N., Okakura, K., Moriya, T., Watanabe, M. and  
AUTHORS  
Streptomyces. Streptomycetaceae; Streptomycetes.  
Bacteroid; Actinobacteriales; Actinomycetales; Streptomyces.  
Streptomycetaceae; Streptomycetes.

TITLE	Malaria
Para-position derivatives of fungal antihelminthic cyclodepsipeptid engineered with <i>Streptomyces venezuelae</i> antibiotic biosynthetic genes	1

PUBMED  
 15184904  
 2 (bases 1 to 5251)  
 Yanaï, K., Sumida, N., Okakura, K., Moriya, T., Watanabe, M. and  
 AUTHORS  
 REFERENCE  
 22 (117, 370-373) (2007)  
 Nucleic Acids Res. 35:117-123 (2007)

**TITLE**  
Direct Submission  
**SUBMITTED** (01-AUG-2003) Koji Yanai, Mei-ji Seika Kaisha, Ltd.,  
Microbiological Resources and Technology Laboratories, 788 Kayama

**FEATURES**

(E-mail: [koji.yana@meiji.co.jp](mailto:koji.yana@meiji.co.jp), Tel: 81-465-37-5106,  
Fax: 81-465-37-6397)

Location/Qualifiers

```
/organism="Streptomyces venezuelae"  
/mol_type="genomic DNA"  
/db_xref="taxon:54571"
```

```

CDS
    /gene="papa"
    651..2711
    /gene="papa"
    /start=1

```

```
/evidence=experimental
/transl_table=1
/product="4-amino-4-deoxychorismate synthase"
/protein_id="WP021140.1"
```

```

/db_xref="GI:47846866"
/translation="MRTLLIDNYSFTNLFQYIGKATGQPPVVPNDADWSRLPV
FAIVSPGPGSPDRERDGISSRAITDSGLPYLVGLGHGIIAQLFGGTGLAPE
HGRSEVDRHTGDEVERGI.DSDPTAVPVNHS.AATN.DDEL.FPI.AWSDGIVMGI.RHP

```

PLNGVQFPIHESISDFGREIMANFRDLALHHRARHGADSPYELHVRVDVLPDA



```

VRGCLPGEITFMDSSVLEBGRSPFLGDDRGPLAEYLTVRVADGVSVKSGDGT
TTTRRPFVYLEEOLERRRVPAPLEPFEFNIGYVLYGELEKAEKRTGPAPRSPH
DAFLPADRAIALDHOEGCCYLALDRBGHDGARMLETAETALRTGAPARAPAP
AMVFGIPEAAGRGPLARAHDDAYIKRIDECCKEKNESYEICITNVYTPTEAT
ALPLYSALRAISFPYIGALFPELVSLSPERFLTGADGVESPIKGTAPRGST
AEEDERLRADLAGREKDAENLMIVDLRNDLSVCAIGVHVRLEVEVTYAPVOL
VSTIRGLRPGSTTACRAAPFGSGMTGPKKRTWMIIDIRLEGGPGVSGALGMFA
LSGADLSIVIRITVLADGQAFRGVGAIVSLSDQEEFETVTKARAVTALDSAV
AGAR"
gene
2708. .3676
/ gene="papC"
CDS
2708. .3676
/ gene="papC"
/ codon_start=1
/ evidence=experimental
/ transl_table=11
/ product="4-amino-4-deoxyphenate dehydrogenase"
/ protein_id="BAD21141.1"
/ db_xref="GI:47846867"
/ translation="MSGFPRSVVVGSGAGVGMFAGLREAGSRTLVVDLPVPRGD
ACLVGVTAARGPELAALRADLVLAVHEDVALKVAPTVTRMRPGALLADLSVST
GMAELAAHPGVQHYGNMPMPAPAGMTGPVAAVTRPGVTTALLRLVEGGGRP
VRLTAEHEDRTATQALTHAVILSLFGLARLGVDRALAAAPPHQVLLALLARV
LGSSPEYVGDIGQSNPRAPASRALAEALRSFALVGDDBDADAPGRADAPGHPGGC
DGAGNLDGVGEIRLRLMGPPLAAGODHCQELFRTLHRTDEGEKDR"
3673. .3984
/ gene="papB"
CDS
3673. .3984
/ gene="papB"
/ codon_start=1
/ evidence=experimental
/ transl_table=11
/ product="4-amino-4-deoxychorismate mutase"
/ protein_id="BAD21142.1"
/ db_xref="GI:47846868"
/ translation="MTEONELQRLRAELDALDGLTDTVRRITDLGVRIARYKRGV
PMQPGRVSLVXDRARAYADHGLDESFLVNLVDVITTECRVEDLVMSRESLTAEDR
R"
4065. .5243
/ gene="ORF1V"
CDS
4065. .5243
/ gene="ORF1V"
/ codon_start=1
/ evidence=not experimental
/ transl_table=11
/ product="hypothetical protein"
/ protein_id="BAD21143.1"
/ db_xref="GI:47846869"
/ translation="MAATTTLALTGRTAQQPYPMPDPAVDVTEBELAGLPVLTPEB
EYVDLNMGMALVARGAALVGGDCARFHEAVPDVVRKVDNLQGLAAMRAGSLP
AYALGRIAGQYGRSSPPTESGAKRKPSYCGDAVNEPEFEPKARTPEPRRLTA
YTOSQIVLDEIRKSWGRPVLEKVTYSHELPLPYEPVLRBEAGHGTSGSAHFHMG
ERTREBDGAVALAQAQVHNVPVGLPTVSPEDAVLSRLNPEGPVGRITFIVRGA
KEVDELPLPVRAVARHGAAPVWMLCDPMHNGSLAGHKTRLIEPMRAETAAVETLR
EHQWPAHGLHLELTPDVTECVSELDPRPFTDYRSTCDPLNPEQSAADVTHFLSL
"
ORIGIN
Query Match 100.0%; Score 312; DB 1; Length 5251;
Best Local Similarity 100.0%; Pred. No. 2e-33;
Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 181 CGCTACGCCCGCAGCACCAGCGCTTCGAGCAATCGTTCGTGTGAACCTTACGACGTATC 240
| | | | |
DB 3853 CGCTACGCCCGCAGCACCAGCGCTTCGAGCAATCGTTCGTGTGAACCTTACGACGTATC 3912
| | | | |
QY 241 ATCAGCAGATGTGCGCGTCGAGGACCTGTGTATGAGCCGGAGAGCCCTGACGCCGAG 300
| | | | |
DB 3913 ATCAGCAGATGTGCGCGTCGAGGACCTGTGTATGAGCCGGAGAGCCCTGACGCCGAG 3972
| | | | |
QY 301 GACCGCGGTGA 312
| | | | |
DB 3973 GACCGCGGTGA 3984
| | | | |
RESULT 4
AF262220 14159 bp DNA linear BCT 09-JAN-2002
LOCUS
DEFINITION
Streptomyces venezuelae chloramphenicol biosynthetic gene cluster,
partial sequence.
ACCESSION
AF262220 GI:14290414
VERSION
AF262220.2
KEYWORDS
SOURCE
ORGANISM
Streptomyces venezuelae
Streptomyces venezuelae
Bacteria; Actinobacteridae; Actinobacteriales;
Streptomycetaceae; Streptomyces.
REFERENCE
1 (bases 1 to 14159)
He,J., Magarvey,N., Pirae,M. and Vining,L.C.
The gene cluster for chloramphenicol biosynthesis in Streptomyces
venezuelae ISP5230 includes novel shikimate pathway homologues and
a monomolecular non-ribosomal peptide synthetase gene
Microbiology 147 (Pt 10), 2817-2829 (2001)
JOURNAL
MEDLINE
21461106
PUBMED
11577160
REFERENCE
2 (bases 1 to 14159)
He,J., Magarvey,N.A. and Vining,L.C.
Direct Submission
JOURNAL
Submitted (01-MAY-2000) Biology, Dalhousie University, 1355 Oxford
St, Halifax, NS B3H 4J1, Canada
REFERENCE
3 (bases 1 to 14159)
He,J., Magarvey,N.A. and Vining,L.C.
Direct Submission
JOURNAL
Submitted (05-JUN-2001) Biology, Dalhousie University, 1355 Oxford
St, Halifax, NS B3H 4J1, Canada
REMARK
COMMENT
Sequence update by submitter
On Jun 5, 2001 this sequence version replaced gi:10716943.
FEATURES
source
1. 14159
/organism="Streptomyces venezuelae"
/mol_type="genomic DNA"
/strain="ISP5230"
/db_xref="taxon:54571"
/map="adjacent to papAB"
233. .3511
/function="essential for chloramphenicol biosynthesis"
/codon_start=1
/transl_table=11
/product="CmIG"
/protein_id="AAG21974.1"
/db_xref="GI:10716944"
/translation="MTAGSPAGRTPLRPALAAVRDHSPLADIGMCDVNPCTAFV
GYVEPRVDDPAAMREFVARKLGLTRALADLVEAEEDVYVTTDPRGPGVGGAEALAD
LVTAYNRSPLKIDEAVRLAYMKHDGEYLVFKRSHLVLADADVDLKQIRAHILRG
DTPPRIGIRYRHARTVQYALPAADAEVTRALGELVPERKGIPTISAVEBWL
PLREGVTFDELISAIVGLVQALGGLVQVYSRWEPAFKGQYVEIPLVVRPELP
SDTPEYPAKTHLAOESLGRFTMSDLDAPARAFTSGRMPRVVVSDDTFMRPEPDRRW
VPRRSARVREDLKFLADRSWPGPPLRMQYKRGFLSPGAADVLDLKKRIASGCTP
"
3498. .6404
/note="non-ribosomal peptide synthetase; contains
reductase-like domains near C-terminus"
/codon_start=1
/transl_table=11
CDS
3498. .6404

```



```
/product="Cm1H"
/protein_id="AAG21975.2"
/db_xref="GI:14327897"
/translation="MELPDLITDLPFRHGRGRTALRTAGRTWTVEELDRVTSALARRID
ACCPARRVIVGEHTAEAVYMAALAMRSHAVTTPNRRGLPARREFEPVADAALIV
CERRRALVNGEKRALYAGDYGMPTDPPADGTDSPASVAVSIFTSIGTD
PKIVDGHGGLNLCSLRRLDITPDQVLAHSIFSPASVELIGTYAGATVIVP
VRDQASVSVSHLAHGCIDLMLSPSVARLDEARSRIRVEFCGALGEGYDK
AAVSRSFNAVYPTVATVCSLAELTSTPSTPTGPFAYRDPDSDGHAATGE
LIVGQVVALGVAGSGSPAEVEFGVTSAGVATGVVSLSDGELTYGRIDEQIKR
LGRVNLAVHGSTSLRHLGREVALVRQDATTLLVTADEGATPESLMARIDIVPWE
APRUVLVDALEPTSGSKDRSLRLSLSPACAPGCTDGEDAELRLVDVTVL
GQETIGETSIIDFGSSSLAMIQIVKLSDAVEAAFAANDVFPAPAFRLHJRG
EAVAPASAVDTLLRRETEREDALRAELRLRDTHEPVPAGADGDEVLTLGASGF
IGHVLDRLLAARPVIVSTGDPDGLTGHAIRFORADYARVRAISYAELEKRW
RRGPVDAVHVGQVNHLLPLDHLSSGVTALVRAAALGARSPAFISAAASG
ADPLISAATLTAVGDPYRSKLSSEYVNTLAVLGCANSHRPGVYGRPERDRHL
KDDWFTALLETARRVAMPRLSGHVPCVDGTLADTLGLGPDANPGTADGASRTDSR
SAVVHRTYALDELRLHGTLEADVLAAPAFERRVKGSEVPAPLLAQAALSGPW
PSAHRVDHDIIGRLITGPTDTPADRPERTGTSRGRERRRRADTALTGEIP"
```

Query Match 95.0%; Score 296.4; DB 1; Length 14159;  
Best Local Similarity 98.4%; Pred. No. 2e-31;  
Matches 311; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

Db 1 ATGACCGAGCAGAAGAGCTGCA----GGCGCTGCGCGCGAGAGCTCGACCGCTCGACGG 56  
11142 ATGACCGAGCAGAAGAGCTGCAAGCTTGGCGCTGCGCGCGAGAGCTCGACCGCTCGACGG 11201

QY 57 GACGCTCTTGAGACAGCGTGGCGCGCGCATGACCTTCGATGTCGCGATCGCGCGGTACAA 116  
Db 11202 GACGCTCTTGAGACAGCGTGGCGCGCGCATGACCTTCGATGTCGCGATCGCGCGGTACAA 11261

QY 117 GTCCCGCAGCAGCGGTCCCATGATGACAGCCCGCGCGGTGACCTGCTCAAGACAGGGC 176  
Db 11262 GTCCCGCAGCAGCGGTCCCATGATGACAGCCCGCGCGGTGACCTGCTCAAGACAGGGC 11321

QY 177 CGCCCGCTAGCGCGCGAGCAGCGGCTCGAGCATGCTTCRGTGTAACCTCTACAGACGT 236  
Db 11322 CGCCCGCTAGCGCGCGAGCAGCGGCTCGAGCATGCTTCRGTGTAACCTCTACAGACGT 11381

QY 237 GATCATCAGCAGATGTCGCGCTCGAGGACCTGTGATGAGCCGGAGACGCTTGAACGCG 296  
Db 11382 GATCATCAGCAGATGTCGCGCTCGAGGACCTGTGATGAGCCGGAGACGCTTGAACGCG 11441

QY 297 CGAGGACCGCGCGGTGA 312  
Db 11442 CGAGGACCGCGCGGTGA 11457

ORIGIN

misc\_feature  
/note="Region: adenylation domain A1"  
3708..3746  
/note="Region: adenylation domain A2"  
3339..3386  
/note="Region: adenylation domain A3"  
4089..4100  
/note="Region: adenylation domain A4"  
4350..4370  
/note="Region: adenylation domain A5"  
4494..4538  
/note="Region: adenylation domain A6"  
4593..4610  
/note="Region: adenylation domain A7"  
4647..4706  
/note="Region: adenylation domain A8"  
4833..4853  
/note="Region: adenylation domain A9"  
4893..4910  
/note="Region: adenylation domain A10"  
5043..5075  
/note="Region: thiolation domain"  
5343..5375  
/note="Region: NADH binding site"

```
RESULT 5
LOCUS A48325 387 bp DNA linear PAT 07-MAR-1997
DEFINITION Sequence 3 from Patent WO9601901.
ACCESSION A48325
VERSION A48325.1 GI:2302118
KEYWORDS Streptomyces pristinaespiralis
SOURCE Streptomyces pristinaespiralis
ORGANISM Streptomyces pristinaespiralis
REFERENCE 1 (bases 1 to 387)
AUTHORS Blanc,V., Thibaut,D., Bamas-Jacques,N., Blanche,F., Crouzet,J.,
Barriere,J., Debussche,L., Famechon,A., Paris,J. and
Dutric-Rosset,G.
TITLE STREPTOMYCINASES AND METHOD FOR PREPARING SAME BY MUTASYNTHESIS
JOURNAL Patent: WO 9601901-A 3 25-JAN-1996;
COMMENT RHONE-POULENC RORER SA (FR)
Other publication AU 2891295 960209
Location/Qualifiers
source 1..387
/organism="Streptomyces pristinaespiralis"
/mol_type="unassigned DNA"
/db_xref="taxon:38300"
```

Query Match 43.5%; Score 135.6; DB 6; Length 387;  
Best Local Similarity 69.8%; Pred. No. 2.5e-09;  
Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 14 ACAGCTGAGCGGCTGCGCGCGCGAGCTGAGCCCTCGACGGAGACGCTCTGAGACGG 73  
Db 59 ACCCTTCGACGGGCTGCGCGCGCGCGCTGAGCGCGCGAGCGCGCGCTCTGAGACGG 118

QY 74 TGGGCGCGCATGCACTGCGTGTCCGATCGCGCGGTAAAGTCCCGCACGGCGTCC 133  
Db 119 TCCGACAGCGCTGAGCATCTGCTCGCATGCGGAGTACAAAGCCCTCCACAGTGC 178

QY 134 CGATGATGAGCGCGCGCGCGGTGACCTGTGCAAGACAGGGCGCGCGCTACGCCCGCG 193  
Db 179 CGATGATGAGCGCGCGCGCGCGATGCGCGAGTGCAGCGCAACCGCGCGCTACGCCCGCG 238

QY 194 ACCAGCGCTCGACCGAATGCTTCGTGTAACCTCTACAGCGTATGACAGAGATGT 253  
Db 239 ACCAGCGCATGACCGCGCGCTTCGTGCGACCTGTACAGACGATCATCACGAGACCT 298

QY 254 GCCGCGTCAAGACCTGTGAT 275  
Db 299 GCCGCGTCAAGACGATGAT 320

RESULT 6  
LOCUS AR19835 387 bp DNA linear PAT 20-APR-2002  
DEFINITION Sequence 4 from patent US 6352839.  
ACCESSION AR19835  
VERSION AR19835.1 GI:20248204  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 387)  
AUTHORS Blanc,V., Thibaut,D., Bamas-Jacques,N., Blanche,F., Crouzet,J.,  
Barriere,J.-C., Debussche,L., Famechon,A., Paris,J.-M. and  
Dutric-Rosset,G.  
TITLE Streptogramins for preparing same by mutasynthesis  
JOURNAL Patent: US 6352839-A 4 05-MAR-2002;  
FEATUERS Location/Qualifiers  
source 1..387  
/organism="unknown"



ORIGIN /mol\_type="unassigned DNA"

Query Match 43.5%; Score 135.6; DB 6; Length 387;  
Best Local Similarity 69.8%; Pred. No. 2.5e-09;  
Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 14 ACAGAGCTGCAGCGGCTGCGCGGAGCTCGACGCGCTTCGACGCGGACGCTCTTGAGACAGG 73  
DB 59 ACCCCCTCGACGCGGCTGCGCGGCGCCGCTTGAGACGCGCGGAGCGCCGCTTCTGAGACCGG 118  
QY 74 TCGGCGCGCGCATCGACTGCTGCTGCGCATTCGCGCGTCAAGTCCCGGACGCGCGTCC 133  
DB 119 TCGCACACGCGCTGAGCATCTGCTGCGCATTCGCGCGTCAAGCGCTTCACAGGCTGC 178  
QY 134 CGATGATGACGCGCGCGCGGCTCAGCTGCTCAAGACAGGCGCGCGCTTCAGCGCGCG 193  
DB 179 CGATGATGACGCGCGCGCGGATCGCCAGGTCCACGCGCGCGCGCGCTTCAGCGCGCG 238  
QY 194 ACCACGCGCTCGACGATCGTCTGCTGTAACCTCTACGACGATGATCATCGAGAGTGT 253  
DB 239 ACCACGCGCATCGACCGCGCTTCTCTGCGACCTCTGACAGATCATCATCGAGACCT 298  
QY 254 GCCGCGTCGAGACCTGTGTAT 275  
DB 299 GCCGCTCGAGGACGAGTGTAT 320

## RESULT 7

A48323 2888 bp DNA linear PAT 07-MAR-1997  
LOCUS Sequence 1 from Patent WO9601901.  
DEFINITION A48323  
ACCESSION A48323  
VERSION A48323.1 GI:2302116

## KEYWORDS

Streptomyces pristinaespiralis  
Streptomyces pristinamycinis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

## REFERENCE

1 (bases 1 to 2888)  
Blanc, V., Thibaut, D., Bamas-Jacques, N., Blanche, F., Crouzet, J.,  
Barriere, J., Debussche, L., Pamechon, A., Paris, J., and  
Dutric-Rosset, G.

STREPTOMYCINES AND METHOD FOR PREPARING SAME BY MUTASYNTHESIS  
Patent: WO 9601901-A 1 25-JAN-1996;

RHONE-POULENC KOREA SA (FR)  
Other publication AU 2891295 960209  
Other publication FR 2722210 960112.

FEATURES  
Location/Qualifiers

1..2888  
/organism="Streptomyces pristinaespiralis"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:38300"

## ORIGIN

Query Match 43.5%; Score 135.6; DB 6; Length 2888;  
Best Local Similarity 69.8%; Pred. No. 1.5e-09;  
Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 14 ACAGAGCTGCAGCGGCTGCGCGGAGCTCGACGCGCTTCGACGCGGACGCTCTTGAGACAGG 73  
DB 1931 ACCCCCTCGACGCGGCTGCGCGGCGCCGCTTGAGACGCGCGGAGCGCCGCTTCTGAGACCGG 1990  
QY 74 TCGGCGCGCGCATCGACTGCTGCTGCGCATTCGCGCGTCAAGTCCCGGACGCGCGTCC 133  
DB 1991 TCGCACACGCGCTGAGCATCTGCTGCGCATTCGCGCGTCAAGCGCTTCACAGGATGC 2050  
QY 134 CGATGATGACGCGCGCGCGGCTCAGCTGCTCAAGACAGGCGCGCGCTTCAGCGCGCG 193  
DB 2051 CGATGATGACGCGCGCGCGGATCGCCAGGTCCACGCGCGCGCGCTTCAGCGCGCG 2110  
QY 194 ACCACGCGCTCGACGATCGTCTGCTGTAACCTCTACGACGATGATCATCATCGAGAGTGT 253

DB 2111 ACCACGCGCATCGACCGCGCTTCTCTGCGACGCTCTGTAAGACAGATCATCATCGAGACCT 2170  
QY 254 GCCGCGTCGAGACCTGTGTAT 275  
DB 2171 GCCGCTCGAGGACGAGTGTAT 2192

RESULT 8  
AR198353 2888 bp DNA linear PAT 20-APR-2002  
LOCUS Sequence 1 from patent US 6352839.  
DEFINITION AR198353  
ACCESSION AR198353  
VERSION AR198353.1 GI:20248202

KEYWORDS  
Unknown.  
ORGANISM  
Unclassified.  
REFERENCE  
1 (bases 1 to 2888)  
Blanc, V., Thibaut, D., Bamas-Jacques, N., Blanche, F., Crouzet, J.,  
Barriere, J.-C., Debussche, L., Pamechon, A., Paris, J.-M. and  
Dutric-Rosset, G.

Streptogramins for preparing same by mutasynthesis  
Patent: US 6352839-A 1 05-MAR-2002;  
Location/Qualifiers

1..2888  
/organism="unknown"  
/mol\_type="unassigned DNA"

## ORIGIN

Query Match 43.5%; Score 135.6; DB 6; Length 2888;  
Best Local Similarity 69.8%; Pred. No. 1.5e-09;  
Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 14 ACAGAGCTGCAGCGGCTGCGCGGAGCTCGACGCGCTTCGACGCGGACGCTCTTGAGACAGG 73  
DB 1931 ACCCCCTCGACGCGGCTGCGCGGCGCCGCTTGAGACGCGCGGAGCGCCGCTTCTGAGACCGG 1990  
QY 74 TCGGCGCGCGCATCGACTGCTGCTGCGCATTCGCGCGTCAAGTCCCGGACGCGCGTCC 133  
DB 1991 TCGCACACGCGCTGAGCATCTGCTGCGCATTCGCGCGTCAAGCGCTTCACAGGATGC 2050  
QY 134 CGATGATGACGCGCGCGCGGCTCAGCTGCTCAAGACAGGCGCGCGCTTCAGCGCGCG 193  
DB 2051 CGATGATGACGCGCGCGCGGATCGCCAGGTCCACGCGCGCGCGCTTCAGCGCGCG 2110  
QY 194 ACCACGCGCTCGACGATCGTCTGCTGTAACCTCTACGACGATGATCATCATCGAGAGTGT 253  
DB 2111 ACCACGCGCATCGACCGCGCTTCTCTGCGACGCTCTGTAAGACAGATCATCATCGAGACCT 2170  
QY 254 GCCGCGTCGAGACCTGTGTAT 275  
DB 2171 GCCGCTCGAGGACGAGTGTAT 2192

## RESULT 9

SPU60417 4740 bp DNA linear BCT 07-MAR-1997  
LOCUS SPU60417  
DEFINITION Streptomyces pristinaespiralis 4-dimethylamino-L-phenylalanine  
precursor biosynthesis (papA, papC, papB, papM) genes, complete  
cds.  
VERSION U60417.1 GI:1575335

KEYWORDS  
Streptomyces pristinaespiralis  
Streptomyces pristinamycinis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE  
1 (bases 1 to 4740)  
Blanc, V., Gil, P., Bamas-Jacques, N., Lorenzon, S., Zagorec, M.,  
Thibaut, J., Bisch, D., Blanche, F., Debussche, L., Crouzet, J. and  
Schleutinger, J.

Identification and analysis of genes from Streptomyces  
pristinaespiralis encoding enzymes involved in the biosynthesis of



JOURNAL Mol. Microbiol. 23 (2), 191-202 (1997)  
MEDLINE 97197164  
PUBMED 5044253  
REFERENCE 2 (bases 1 to 4740)  
AUTHORS Blanc, V., Gil, P., Bamas-Jacques, N., Lorenzon, S., Schlemminger, J., Bisch, D., Blanche, F., Debussche, L., Crouzet, J., and Thibaut, D.  
TITLE Direct Submission  
JOURNAL Submitted (11-JUN-1996) Recherche Pharmaceutique, Rhone-Poulenc Rorer, 13 quai Jules Guesde, B. P. 14, Vitry sur Seine cedex 94403, France

FEATURES  
source Location/Qualifiers  
1..4740  
/organism="Streptomyces pristinaespiralis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:38300"  
68..2227  
/gene="papa"  
68..2227  
/gene="papa"  
/function="p-aminobenzoate synthase"  
/codon\_start=1  
/transl\_table=11  
/product="PAPA"  
/protein\_id="AAC44866.1"  
/db\_xref="GI:1575336"  
/translation="MRTVRLIDNYDSFTYNYLFQMLAEVNGAPLVVRNDRTWQA  
LAPGPDNVVSPGPDATDGLSRVITWMDPLGVGCHGALCLAAAVH  
APBPHGRSDIRHDOGGLPANTPSPLTVRHSILVRQLPADLRPAHRAQOIMAV  
AHRHLPRFGVQHPBESTSEHGRMLANFRDLSRAAGHPPTHERIPAPAPAPAP  
APAPAPAPVGEYRLHREYACVPDADAAFTALFADAPAFWLDSSRVEGLARFTFL  
GAPAGPLGEQITVDADRAVAVDGGGGERRPGTLFDLHEHLARALPATGLPEEF  
NLGVYLYGETEKADSGEDAHRGELPDGAFMADRLALDHEGARMALALASSTER  
ATAPABRWLTDATLATTAPRPFLLPDDQLPADVHYRSLPRVYLVEGRL  
ITDGETYEVCLTMLRVPGRIDPLTAYRALRTVSPA VYAYIQPGATYLSSEERL  
RIGADMABESKPIKTRPRGAGAPDAAYVAGSAAABKRSSEMLMVDLYRNDLGC  
DISGVVPGLEFEVETATVHQLSTVGRILAADVSRRVRAAFPGSGMTGAKVATM  
QFTDLREKGRGVSGALGVALSGAADLSIVRTIVATEEAATIGVGAVALSDPD  
DEVREMLIKQITLALRQAAGATASDELLAGLSR"  
/complement(2486..3376)  
/gene="papC"  
complement(2486..3376)  
/gene="papC"  
/function="dehydrogenase"  
/codon\_start=1  
/transl\_table=11  
/product="PAPC"  
/protein\_id="AAC44867.1"  
/db\_xref="GI:1575337"  
/translation="MRGGSYFGRCVVVGAGAVGRNFSHMLVRSVAVTWLDVAGAGA  
ADGVRVAGDVRPRGPEAVNALAADVVVLAVEPVAWEAVEVLAVMRGAVIADTL  
SVKSRVLAGRLREAPGLOAVGNPMFAPSLGGRVAAVAVVDPGVALVADTL  
GARVEMPARRHDELTAQOAAHTAAVLAFCGLISLADVGAIRPSAPPHAMAL  
LARIAGTPEVYTDIOANFGAPAAQALGRGLVRIGQAVRDEETFAALFAELKGV  
LGHGAELERLCRMFALH"  
3413..3802  
/gene="papB"  
3413..3802  
/gene="papB"  
/function="mutase"  
/codon\_start=1  
/transl\_table=11  
/product="PAPB"  
/protein\_id="AAC44868.1"  
/db\_xref="GI:1575338"  
/translation="MTPPAIPAPPAATGPAATDPLDALRARDAAADALLAVRRL  
EDMLRIGERYRLHQVPMQPHRIAQVNAANAARAADHGDIDPAFLRLTYTITETRL  
BECWIRASGAPVPTPVHASASARGAVS"  
3799..4677  
/gene="papM"  
3799..4677  
/gene="papM"  
/function="N-methylase"  
gene  
CDS

/codon\_start=1  
/transl\_table=11  
/product="PAPM"  
/protein\_id="AAC44869.1"  
/db\_xref="GI:1575339"  
/translation="MTAAPFTLAQALDEATGQITGACTRADARADTRLLAHACQVA  
PGDLDTCLAGPAPPRFMHYVRRRLTEPERERVGAHYEMGHRDLPGVFPKPEEE  
ITRDALFARLEARVREGTAPLVVDLCAGGTWAVTJARVAPARVAGIELSQAAARA  
RNRAGTGARIYOGDARDAPELSGTVDLVVTNPPYIPGLRTASAEVLEHDPPLAM  
AGEEGIGMIKRAMERTARLLAPDGVLLHSGVQLASVAPALFRATGRWHSRRPTCN  
DGLTAVRNRHTCAPA"

ORIGIN

Query Match 43.5%; Score 135.6; DB 1; Length 4740;  
Best Local Similarity 69.8%; Pred. No. 1.3e-09;  
Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 14 ACGAGCTGACAGCGCGCTGCGCGGAGACTGACGCCCTCGACGGAGCGCTCTGACACGG 73  
Db 3471 ACCCCCTCGACGCGCTGCGCGCGCGCTGACGCCCGCGACGCCCGCTCTGACACGG 3530  
QY 74 TGGCGGCGCGCATCGACTCGGTGCGCATCGCGGCTACAAAGTCCCGGACAGGGGTC 133  
Db 3531 TCCGACACGCGCTGACATCTGCTCGCATCGCGGCTACAAAGCGCTTCCACACAGGTGC 3590  
QY 134 CGATGTGACAGCGCGCGCGGCTGCGCTGTCACAGACAGAGCGCGCGCTACCGCGCG 193  
Db 3591 CGATGTGACAGCGCGCGCGGCTGCGCGGCTGCGCGGCTGCGCGCGCGCGCGCGCG 3650  
QY 194 ACCACGCGCTCGACGAATGCTTCTGTGACACTCTACGACGTATCATCAGGAGATGT 253  
Db 3651 ACCACGCGCATCGACCGCGCGCTTCTGCGCACCTGTACGACAGATCATCAGGAGACT 3710  
QY 254 GCCGCGTCGACGACGCTGTGAT 275  
Db 3711 GCCGCTCGAGGACGATGAT 3732

RESULT 10

AF263011S2/c 10029 bp DNA linear BCT 07-JUN-2000  
LOCUS AF263011S2  
DEFINITION Streptomyces griseus subsp. griseus macroretrolyde biosynthesis  
gene cluster, downstream region.  
ACCESSION AF263012  
VERSION AF263012.1 GI:8896099  
KEYWORDS  
SEGMENT 2 of 2  
SOURCE Streptomyces griseus subsp. griseus  
ORGANISM Streptomyces griseus subsp. griseus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycinae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 10029)  
AUTHORS Smith, W. C., Xiang, L. and Shen, B.  
TITLE Genetic localization and molecular characterization of the nons  
gene required for macroretrolyde biosynthesis in Streptomyces  
griseus DSM40695  
JOURNAL Antimicrob. Agents Chemother. 44 (7), 1809-1817 (2000)  
MEDLINE 20316791  
PUBMED 10858335  
REFERENCE 2 (bases 1 to 10029)  
AUTHORS Smith, W. C., Xiang, L. and Shen, B.  
TITLE Direct Submission  
JOURNAL Submitted (02-MAY-2000) Chemistry, University of California at  
Davis, One Shields Ave., Davis, CA 95616, USA

FEATURES

source Location/Qualifiers  
1..10029  
/organism="Streptomyces griseus subsp. griseus"  
/mol\_type="genomic DNA"  
/strain="DSM40695"  
/sub\_species="griseus"  
/db\_xref="taxon:67263"  
205..990  
/gene="nonH"



CDS  
205.. .990  
/gene="nonH"  
/note="similar to fixa"  
/codon\_start=1  
/transl\_table=11  
/product="flavoprotein reductase nonH"  
/protein\_id="AA01240.1"  
/db\_xref="GI:8896112"  
/translation="MSRIYCVKRYVDATGDRRADTLDEEDVGLSEDEVAV  
EOLQIADGADLIVTVGPEDAKALKALSMGADKAHVEDDLHSDVWGTS  
VLAAVEKAGYDLVTCMASITDVGMLPALLAERLGPQVTLSEVAVGCVTGR  
LADGTASEQLASLPAAVSVTDGSEAYPSFKGIMAAKKPVESLDDDLDADEVG  
LAGMTVDSATRPARACTIVDEEGEGRLAGLAGQKFI"  
1051..2013  
/note="orf26; similar to fixb"  
/codon\_start=1  
/transl\_table=11  
/product="flavoprotein reductase"  
/protein\_id="AA01241.1"  
/db\_xref="GI:8896113"  
/translation="MAEVLVVDHVGCAVAKPTLELTLLARRVDPVAVVLGAGAGA  
APVDEHGAVKVLTDAPEFADYLVKVKVLAQAAFAVSPVAVLVSSAEGEILAR  
LALRVSGIITDADLEAGDQGVATQAPALASFTTKSRVGPVITVNSAPVPEP  
APAAVAVLPVSEGAATGTKTSRPRESTGRPELTETAVIVSGRGVAGANFPL  
TEALADSLGAAGVSAVDAWGMVPHTSOYGQTKSVSPOLYIASGISGAIQHRAGWQ  
TSKTIYAINKDAAPIFELVYGVGLFAVYQLTPEINTRK"  
complement (2090..3481)  
/note="orf27"  
/codon\_start=1  
/transl\_table=11  
/product="amino acid monooxygenase-like protein"  
/protein\_id="AA01242.1"  
/db\_xref="GI:8896114"  
/translation="MTGPRTPLDLLGVAGPFTSLAALADGVPGLRTAFHEQRAA  
FHHMGLLIGATLOVFLADLVSLVPTSPMSYLNLIKRRRLFPFYAEHFIHHT  
EFNNYLRWSTELPATRFGRVDVTAMDPODAFAVEFTLGDGEGTLTGTHARL  
ALGVGTAPHPEPRLVODPAAVHSAVYLSQRNLLAARHITVVGSGQSAVYL  
DLARSREGAGLTWLTARTPARAMEYSKIGLEQFTDYTRYHGLPQARLDLPRO  
WQILKVGSDTLGDIHDELTRSLDGGMPVLTLPGLIEVGAATEGRVLELVENQ  
QEARGLVTVDAVVLATGYRERPVDLAALDPYIVRESTRPOVDAQRLVLAPELQ  
SVFVONAEHTHGTGTPDLGLAAMRSVAVIYNLTGKEFYLPERTAFTTGLGARDSD  
DRDTASRAEERR"  
complement (3552..5129)  
/note="orf28"  
/codon\_start=1  
/transl\_table=11  
/product="amino acid decarboxylase-like protein"  
/protein\_id="AA01243.1"  
/db\_xref="GI:8896115"  
/translation="MSDGTPTIPGPAPREGVAGAGHGRRLAATLAVLALDEMA  
PARGRLPAGVEVVALSTLARCAPVLEEGVGEALGAVVAAGADPADPACV  
AHLACPLAAVAVADLAGALNPSMWDPAAGVLELTALALHPTAPADPA  
LITSGGTESNLVALLARERAVETLGVANSHVRAAMMLGLPAPVVAACRG  
RIDPALDDALASLAGAPLTVATAGTGGGIDPPEIRAVRERBARLHVAAACG  
PLFSERLAPLAGLEHAASVTDLHGLGMPVAAAGVLAADGMLALSLADYANA  
DDTTEAGLPULGRSITRTTRPDALKAAVTFRALGRGLSELVEHCRTALHGRADV  
TRPGLRRRPDGLISTVLFRTVADALPDAGAAVVAAVRITLLAERAVLGAQED  
GDGRRRLMKATLLHPDAAAELLPLDLIVTAADRAATEADGVVTGADRAATE  
APGAEDTDAEAERLSP"  
complement (5438..6403)  
/note="orf29"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AA01244.1"  
/db\_xref="GI:8896116"  
/translation="MREVTCHVCAQRDLDAALSGIRTMGRHMKRYDDASERLRE  
TABLDLHVAAVADLDDVARSALRTAKCHLGEWSGCFPDGQBLHPFLIGT  
LTSBDIAFGDVVRRGGQAPSAGTWDAFVSVSGVWGMQRTIGLLINDVAPALH  
DGVYSLDASDPTDLAAMDALCPYLAEAGHQPRMPTVPLRRPADDLALAAAL  
DASGALTPDQRLRLVLLDDQAAFEELASRLVGRASVGPDPAPRTLLPLHALALA  
LAVOAHGMEGLSGSYLPPELLGTPLALRRAGSRTNNLGWYAK"  
complement (6526..8850)

ORIGIN  
Query Match  
Best Local Similarity 50.7%; Score 64; DB 1; Length 10029;  
Matches 154; Conservative 0; Mismatches 150; Indels 0; Gaps 0;  
/note="orf30"  
/codon\_start=1  
/transl\_table=11  
/product="5-methyltetrahydropteroylirriglutamate-  
homocysteine methyltransferase-like protein"  
/protein\_id="AA01245.1"  
/db\_xref="GI:8896117"  
/translation="MTAKPAAAARATVYGYPROGPNRLKRAIBGWKGRVDAALR  
ETRAELRRGTWQQLAEGCHDVPTEGPFSTYDHLDTDSWYAGVPGHRAVDDALDG  
YPMAGQTDVAPLEMTKMPDNYHYLVPELGDVFTVDSAKOVAECALALGHTP  
RPLVGPVTVLLAKPAPVADFEPLTLDRLLPYAEVLADLRAAGAMVOLDEPA  
LVDRTVPVLELMAARAYRELGLADRPOLIVASYFGLGALTIVLAKAPDGLALFT  
ESGAGNLDLAAAGGLPGKRLVGVNGVNRINWINDYKSLATGTGLGLADRVASAS  
CSLIHPLDADAERDIDPOLIAWLAFAARQCTEIVYLAAGLTGGTDTIAELANRAD  
LASRAAALTRDPARATVATVADAGRSQPTTESAQAORAHGLPILPTTIGSP  
QTTELTARADLAKRIDAGYHERIKDEIRVLSYQEARGLVLVHGEPRNDWQY  
FAPQLTGYLATQHGWOVSQYTRVPRPVLAGDISREPMVTRWTTVAQSLTAPVKM  
LTPGVITLAMSFVRDQPLGDPARQVALALRDEVDLEASGTSVIOVDEPALRETLPL  
RAEHAAYLEMATSEFRLATGVRPDTQIHTHVCYAEFGDIVOAIDLDDVLSLEAA  
RSHMOYARELAHAGYPRGAGPGVYDIHSRPISTEEAALLRKGLEAIPASRLMVPD  
GILKTGMETTRASLENLVAAARETAAETPEAA"  
4 ACCGAGCAGAACGAGCTGCAGCGCGCTGCAGCGGAGCTGCAGCGGAGCGCTC 63  
|||||  
4631 ACCGAGTCCAACTGTGTGCGCTGTGCTGCCCGGAAACGACCGCGTCCCCGCACTC 4572  
|||||  
64 CTGACACAGGTGGCGGCGCGCGCATCGACTTGGTGTCCGATCGCGCGGTCAAGTCCCGG 123  
|||||  
4571 CGGATGGTACCGGAGCGCAACCGCCACCAAGGCTCCACCGCGCGGTGATGCTCGAT 4512  
|||||  
124 CACGCGCTCCGATGATGACGCGCGCGGCTGAGCTGGTCAAGGACAGGCGCGCGCG 183  
|||||  
4511 CTGCGCGCGCGCGTGTGCTGCTGCGCTGCGGACGCGCGGATCAACCGCGCGCGCTGCAC 4452  
|||||  
184 TACCGCGCGCAGCAGCGCGCTCGACGATCGTTCCTGAGTGAACCTTAAGCATGATCATC 243  
|||||  
4451 GACGCGCTGGCGTCCCTCGCGCGGAGCGCGCTCGTGGTGTGCGCACCGCGCGGAAACACCC 4392  
|||||  
244 ACCGAGATGTGCGCGCTCGACGACCTGTGATGACCGCGGAGAGCTGACGCGCGAGGAC 303  
|||||  
4391 GACGAGGCGCGCATCAACCGCTCCCGAGATGCGCGCGCTCGCGAGCGGCAACCGCGCC 4332  
|||||  
304 CGGC 307  
|||||  
4331 CGAC 4328  
Db  
RESULT 11  
AE005068/c 13635 bp DNA linear BCT 12-FEB-2001  
DEFINITION Halobacterium sp. NRC-1 section 99 of 170 of the complete genome.  
VERSION AE005068.1 GI:10581033  
KEYWORDS  
SOURCE  
ORGANISM Halobacterium sp. NRC-1  
Halobacterium sp. NRC-1  
Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
Halobacteriaceae; Halobacterium.  
1 (bases 1 to 13635)  
REFERENCE  
AUTHORS Ng,M.V., Kennedy,S.P., Mahairas,G.G., Bergquist,B., Pan,M.,  
Shukla,H.D., Iasky,S.R., Balliga,N., Thorason,V., Shropan,J.,  
Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Weitz,R., Geo,Y.A.,  
Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,  
Maddocks,D.G., Jablonicki,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,  
Isenbarger,T.A., Beck,R.F., Pohlischrod,M., Spudich,J.L.,  
Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,  
Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and  
DasSarma,S.



TITLE Genome sequence of Halobacterium species NRC-1  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (22), 12176-12181 (2000)  
MEDLINE 20504483  
PUBMED 11016950  
REFERENCE 2 (bases 1 to 13635)  
AUTHORS Ng, W. V., Kennedy, S. P., Mahairas, G. G., Berquist, B., Pan, M., Shukla, H. D., Lasky, S. R., Baliga, N., Thorsen, V., Sbrogna, U., Swartzell, S., Weir, D., Hall, J., Dahl, T. A., Weitz, R., Geo, Y. A., Leitner, B., Keller, K., Cruz, R., Danson, M. J., Hough, D. W., Maddocks, D. G., Jablonski, P. E., Krebs, M. P., Angevine, C. M., Dale, H., Isenbarger, T. A., Peck, R. F., Polischod, M., Spudich, J. L., Jung, K. H., Alam, M., Freitas, T., Hou, S., Daniels, C. J., Dennis, P. P., Omer, A. D., Ehardt, H., Lowe, T. W., Liang, P., Riley, M., Hood, L. and Dassarma, S.  
TITLE Direct Submision  
JOURNAL Submitted (14-JUL-2000) Institute for Systems Biology, 4225 Roosevelt Way NE, Seattle, WA 98105, USA  
FEATURES  
source Location/Qualifiers  
1. 13635  
/organism="Halobacterium sp. NRC-1"  
/mol\_type="genomic DNA"  
/strain="NRC-1"  
/db\_xref="taxon:64091"  
complement(142..633)  
/gene="VNG1547C"  
complement(142..633)  
/gene="VNG1547C"  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="Vng1547C"  
/protein\_id="AAG19830.1"  
/db\_xref="GI:10581034"  
/translation="MDRVSKASPRIRAYGVDELNAVGAIRPTDIADLDWLREXON  
HHVVQADNPPEPDDPOTAEHTVHCHEARLDPLESFIILGGSEAGAL  
HARAVCRRAERAAVALQDHPDANDRAVAIYNLSDALVFAALADVAGVHES  
PSY"  
complement(681..920)  
/gene="VNG1546H"  
complement(681..920)  
/gene="VNG1546H"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="Vng1546H"  
/protein\_id="AAG19831.1"  
/db\_xref="GI:10581035"  
/translation="MTLTLYALDGGCPDCESVITDLAADDIDHETVHVDAHRAAKV  
RASGQRSVVYLDDBRGVVMADSQRIQTAAVTLA"  
complement(1776..2588)  
/gene="VNG1548C"  
complement(1776..2588)  
/gene="VNG1548C"  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="Vng1548C"  
/protein\_id="AAG19832.1"  
/db\_xref="GI:10581036"  
/translation="WYESRAGDLGVRKRNATRIQIIIVETAEROPVSCOETADEI  
GYTAQVSDYLDLVEEGHARKGRVYVTKGVWMLISQTDLRTFVEYSESLE  
EYVDVTAVADADIDDERVLSMDHGVMAATPGADGSAVAVATGAACODGTAEFD  
GMIDYEPGEVTLSTFVVODGSGALPDDAVTAFLAHDIVAAAGVLAIVAVRDGSE  
PIIRFCTASVAPAASRGSLVLLVTVADRLSHTDPLERNLGYEVVGGLD"  
2727..3290  
/gene="cbiI"  
/note="synonym: VNG1550G"  
2727..3290  
/gene="cbiI"  
/note="cbiI"  
/codon\_start=1  
/transl\_table=11  
/product="cobalamin biosynthesis"

gene  
CDS  
/protein\_id="AAG19833.1"  
/db\_xref="GI:10581037"  
/translation="MAQIQLPHDAKAGPTKEPRAVVRSKLATLTDHDFEVSGCTGA  
VTIAARDAGRVTAERKADRLERTKRLAANGADAVELRAAEAPAEI.PADADLFI  
GSRVVDATLDHAETGVDRIYMNVSRLVEVNAAGATQAFDRDRIILEVVQFQVSHGEL  
AAATSPDSNPFYMLVGSASGTEADR"  
3287..3991  
/gene="cbiI"  
/note="synonym: VNG1551G"  
3287..3991  
/gene="cbiI"  
/note="cbiI"  
/codon\_start=1  
/transl\_table=11  
/product="cobalamin biosynthesis"  
/protein\_id="AAG19834.1"  
/db\_xref="GI:10581038"  
/translation="MTVVGIGPSPEDLLTLRGKRTLEADVYVSPGRSLRTVTH  
VDSRIIGDLPFMTSDPEKRLAAMDAAEVAPRARDGVAFATIGDPVYSTFGLR  
RTLDAFHPDIDQTVGVSAVAFTTALGVEVDAGAGALREAAQASPTGDMILF  
KVTDAPTHGELVDAGVTVYGRRLYMEGDTTVTRDAEIAADRBYTLAYERDTLQ  
KDRATAPDTEVAGDA"  
3984..4850  
/gene="cbiF"  
/note="synonym: VNG1553G"  
3984..4850  
/gene="cbiF"  
/note="cbiF"  
/codon\_start=1  
/transl\_table=11  
/product="cobalamin biosynthesis"  
/protein\_id="AAG19835.1"  
/db\_xref="GI:10581039"  
/translation="MPDATQDADDAASQORGRDARVYDHTAGDPTQAGI.PFIGAG  
BDPRLTLVLRHSGDPAIYGAALOMALREBQPTTYLPGVTSFASATLRLQTL  
DAHEGDTVVRHSGDPAIYGAALOMALREBQPTTYLPGVTSFASATLRLQTL  
LNGIANHVAIFTRPQGTLDABEDDIAEFVEMGVDTTCIYGTALPVMERLLAAGD  
PETPTVTVYHASMVPDEDVIEGTIETIGDREVDAGYRASALVIGDAGAGYERSLY  
GDMANRGSDDQ"  
4847..5824  
/gene="cbiG"  
/note="synonym: VNG1554G"  
4847..5824  
/gene="cbiG"  
/note="cbiG"  
/codon\_start=1  
/transl\_table=11  
/product="cobalamin biosynthesis"  
/protein\_id="AAG19836.1"  
/db\_xref="GI:10581040"  
/translation="MSTNDTETDSDGCGSTPDSGVAEDIAIVAFERRKRTAAEI  
VAGIGDRYSDIILDYHGDVFAEHGVDGCVGLMASGIMARKTGLLDDKVEDPAVY  
VYDEALTWAIPLITGSHRANOVADDLASMGAVPANTTSAAGKGVBERAKALDTHV  
VNGDSTVATNLVNDENLGPVARLDGPAVVGVDTVLKRNISDGVVLGTSGVAD  
EQQFRDAMAASLAATDYDIDVEFVATGTRADEGLLAAAEALGLGVVAPEKTLLEA  
HEGPTPSRSKELIGWPIAESAIAIGRDNELLAEKERYDEAVVAVGK"  
5821..6651  
/gene="cobH"  
/note="synonym: VNG1555G"  
5821..6651  
/gene="cobH"  
/note="cobH"  
/codon\_start=1  
/transl\_table=11  
/product="cobalamin biosynthesis"  
/protein\_id="AAG19837.1"  
/db\_xref="GI:10581041"  
/translation="MSDANBEADYGTLLVVGIGPLPGAMTORAKOVVRTADCVVANS  
LYOFLRRDGTLLPADAGERPDQEVRSRTMGRIELAEAFERYVSDGVAHVSGDP  
NVYGSGLVYLAEBEBAAYDVPIELVPGVTAALGAALGAPLSNDPCTISLSDKRG  
WBIIEKURAAIISGFVVLVYCMWRDYSRAIGVLEERADVDVPAIINDAGDGRN  
IDGETHTITTLGSVTDSHGEGMGASLLIGIDETAVMENDHKKLLVTPRGGRDVEDF"



gene	6651..7658	/gene="cbiH"	/note="synonym: VNG1557G"
CDS	6651..7658	/note="CbiH"	/note="CbiH"
		/codon_start=1	/product="cobalamin biosynthesis"
		/protein_id="AAG19838.1"	/db_xref="GI:10581042"
		/translation="MSTDDPTTDDSTNSCGASTTDDSTSCGSGSTTDDASAC	/translation="MSTDDPTTDDSTNSCGASTTDDSTSCGSGSTTDDASAC
		CGASGSDSSCGGASGSDPTDNEQTSATVDDPDLAVGAGPGRPGMTRARA	CGASGSDSSCGGASGSDPTDNEQTSATVDDPDLAVGAGPGRPGMTRARA
		ALDAEHIYGVYTYVLDLPEITDDAEIYDTPWCSEVSTEEARALAGNVAITG	ALDAEHIYGVYTYVLDLPEITDDAEIYDTPWCSEVSTEEARALAGNVAITG
		SGDPNYALAGLALETRESGATPSTMDFEVYGVPAAGSCGARLAPVNDYSISL	SGDPNYALAGLALETRESGATPSTMDFEVYGVPAAGSCGARLAPVNDYSISL
		SDHLTPMPELESRLHAAVEFTISITNPSRKRENFQACEILLAHRAADPVGIV	SDHLTPMPELESRLHAAVEFTISITNPSRKRENFQACEILLAHRAADPVGIV
		HGAGDEDEVEIITTLGLESYGETIDLIDMTTIVIGNAETVYVDDRMVTPRGYETKXD	HGAGDEDEVEIITTLGLESYGETIDLIDMTTIVIGNAETVYVDDRMVTPRGYETKXD
gene	7658..7906	/gene="VNG1558H"	
CDS	7658..7906	/note="VNG1558H"	/note="hypothetical protein"
		/codon_start=1	/product="Vng1558h"
		/protein_id="AAG19839.1"	/db_xref="GI:10581043"
		/translation="MYRVTIDREACDGVFACLTFRDSRFVEDDAGLATLADDTVERT	/translation="MYRVTIDREACDGVFACLTFRDSRFVEDDAGLATLADDTVERT
		KETVYGEFDDDERVADAEDAAACPFPAISVDVAEAD"	KETVYGEFDDDERVADAEDAAACPFPAISVDVAEAD"
gene	7903..8601	/gene="VNG1559H"	
CDS	7903..8601	/note="VNG1559H"	/note="hypothetical protein"
Query Match	19.5% Score 60.8; DB 1; Length 13635;		
Best Local Similarity	51.5% Pred. No. 13; Matches 140; Conservative 0; Mismatches 132; Indels 0; Gaps 0;		
QY	35 CGAGACTCGACGCCCTCGACGGACGCTCTGTGACACGCTGCGCGCGCATCGACCTCG 94		
DB	545 CGGACATCGCGACCTCGACGCGCTCGCGGAGGACAGAACCACTCCAGCTGTGC 466		
QY	95 GTGTCCGATCGCGCGGTACAGTCCCGGACGCGCTCCGATGATCAGCCGCGGG 154		
DB	485 AGGCGGACTTCGCGAACCCCGAGCGCCGACGACCAAGATCACCGCCGACACG 426		
QY	155 TCAGCCGTCAAGGACGAGCGCGCGCTACCGCCGACCAAGCGCTCGACGAATCGT 214		
DB	425 TCACCCAGCTGAGAAATGATTCACCACTGCGAGCGCGCTGAGACCCCTGAGTGTG 366		
QY	215 TCCTGTGAACCTTACGATGATTCATCAAGAGATGTCCCGCTCGAGACCTGTGTA 274		
DB	365 TCATCTCGCCCGGCGGACGAGCGCGCGCGCTGATCAGCCGCGGTGTGCC 306		
QY	275 TGAAGCCGAGAGAGCTTACGCGCCGAGGACCGG 306		
DB	305 GCGCGCGGAGCGCGCGGTGCGCTCCAG 274		
RESULT 12			
AK121213	1964 bp mRNA linear	PLN 29-OCT-2003	
LOCUS	AK121213		
DEFINITION	Oryza sativa (japonica cultivar-group) cDNA clone:J023089K20, full		
ACCESSION	AK121213		
VERSION	AK121213.1	GI:37990836	
KEYWORDS	Full-CDNA; CAP trapper.		
SOURCE	Oryza sativa (japonica cultivar-group)		
ORGANISM	Oryza sativa (japonica cultivar-group)		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		

REFERENCE	1	Enhardtoidae; Oryzae; Oryza.
AUTHORS		
		The Rice Full-length cDNA Consortium, National Institute of
		Agricultural Sciences Rice Full-length cDNA Project Team,
		Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
		Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,
		Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,
		Ohtsuki, K., Shishiki, T., Foundation of Advancement of International
		Science Genome Sequencing & Analysis Group, Otsu, Y., Murakami, K.,
		Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
		Kurosawa, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,
		Nakamura, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Nikura, J.,
		Ikedo, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,
		Kusunegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN:
		Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,
		Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,
		Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Otsu, N., Ota, Y.,
		Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
		Yoshino, M., and Hayashizaki, Y.
TITLE		Collection, mapping, and annotation of over 28,000 cDNA clones from
JOURNAL		Japonica rice
MEDLINE		Science 301 (5631), 376-379 (2003)
PUBMED		22752273
REFERENCE		12869764
AUTHORS		
		Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,
		Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,
		Hayashizaki, K., Hayashizaki, Y., Hayatsu, N., Hiyamato, K., Hirooka, T.,
		Hori, F., Hotta, I., Iida, Y., Iida, Y., Ikeda, R., Imamura, K.,
		Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,
		Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,
		Kikuchi, S., Kishikawa, Hirozane, T., Kishimoto, N., Kobayashi, M.,
		Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
		Koya, S., Kurihara, C., Kurosaki, T., Kusunegi, T., Li, C., Lu, M.,
		Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
		Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, Y.,
		Nakamura, M., Namiki, T., Narikawa, R., Nikura, J., Nishi, K.,
		Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M.,
		Ooka, H., Otsu, N., Ota, Y., Otsu, Y., Otsu, R., Saito, H., Sakai, C.,
		Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K.,
		Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y.,
		Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M.,
		Tagami, Takada, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
		Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K.,
		Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J.,
		Yokomizo, S., and Yoshimura, A.
TITLE		Collection, mapping, and annotation of 28k full-length cDNA clones
JOURNAL		from japonica rice
REFERENCE		Unpublished
AUTHORS		3 (bases 1 to 1964)
TITLE		Kikuchi, S.
JOURNAL		Direct Submission
		Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of
		Agricultural Sciences, Department of Molecular Genetics, Head of
		Laboratory of Gene Expression, 2-1-2 Kamondai, Tsukuba, Ibaraki
		305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp,
		Tel:81-29-838-7007, Fax:81-29-838-7007)
COMMENT		This clone is one of the 32k full-length cDNA clones from japonica
		rice.
		URL: <a href="http://cdna01.dna.affrc.go.jp/cdna/">http://cdna01.dna.affrc.go.jp/cdna/</a>
		NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,
		Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,
		Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,
		Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T.,
		Yamamoto, M., and Nakamura, Y.
		PAIS Genome Sequencing & Analysis Group: Otsu, Y., Iida, Y.,
		Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M.,
		Kobayashi, M., Kodama, T., Kurosaki, T., Kusunegi, T., Lu, M.,
		Masuda, H., Miura, J., Mizuno, K., Nakikawa, R., Oka, M.,
		Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M.,
		Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K., and Murakami, K.
		Genome Exploration Research Group in Riken Genomic Sciences Center
		and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,







```
/gene="VNG1244C"
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1244c"
/protein_id="AAG19604.1"
/db_xref="GI:10580772"
/translation="MTLDLRAEIESIDREIVDLIARTVVAETIADVKAEGSLSTD
ESQEDAMERAGENAAAFQVDNIVKAVFRLILINKIEQRETR"
complement(13877..4731)
/gene="VNG1245C"
complement(13877..4731)
/note="VNG1245C"
/gene="VNG1245C"
complement(13877..4731)
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1245c"
/protein_id="AAG19605.1"
/db_xref="GI:10580773"
/translation="MDGRAVAPAGTVLNLATGVGAAPALDIDTEASVSTPESGV
SGRIAGHPADTALVERCVRYIDRYGDCGGHRTSEVPLAAGLSSAAANATYL
ATDLAGVADVDVADARIGVQARADAGTVTGAFDALAAKMGVAMTDNEDDL
FRAVEMHAAVMTPPERAYADADVARCERSGLAEHVAALAAAGDYATMTVGLAF
CAALDPPTAPAVTALPHAAGVSLSGTSPYVAVGDEBDGIEBSTRMHENGTVAETTT
OLAGARTT"
4819..5247
/gene="VNG1246H"
4819..5247
/note="VNG1246H"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1246h"
/protein_id="AAG19606.1"
/db_xref="GI:10580774"
/translation="MSARNDIAPSTLCEVLAEEGVVEYTDGREAFYNGVPEPIETL
HTNKGQVHLVTDPTREGLVYVNDLKTGHEIDLESTGGRLLDGEBELPQVT
VQADGVAIDIDBETARGVFVFAEDBELSEYAEIYQPA"
complement(15405..6064)
/gene="Ybhf"
/note="synonym: VNG1247G"
complement(15405..6064)
/gene="Ybhf"
/note="Ybhf"
/codon_start=1
/transl_table=11
/product="ABC-type transport protein"
/protein_id="AAG19607.1"
/db_xref="GI:10580775"
/translation="MITATDITKTYGETTALAASVYAMDSGLHCLAPNSGKTTLLR
VLAGLTRHDSGSVTPESLGVAFQTPVNVADLVANLVDVFGALAGATDWRASLVQ
LRIDPVAGRTAGALSGGFOKLDLALALRPAALLDLDEPLADDPATRRRLVDTVTR
YSASHTVVASHTHLDAFADYATLTVVRDGLLDSWDRPAPAPNDPSGVYRLDLRENG
P"
complement(6067..7137)
/gene="VNG1249C"
complement(6067..7137)
/gene="VNG1249C"
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1249c"
/protein_id="AAG19608.1"
/db_xref="GI:10580776"
/translation="MDVSLGKELRWGRHKLALAAVFLVPAVAVTVTGFOHLP
TDIAVAPRTAAVTGDDLRYABAAVTFSPVYGRADLTALGREOVYAAVSVRP
ALADSAGAAFTYIDGSVYPHPSRAIYRLQSLGTRAPDYTSRBEVTOYS
LSTYVSPVSLTAVLALGYLPOTFEAEADRLKRETSLIAVVASKLFWPRLALV
VPLVFPADVAVTAVGGSVLGAVATLAVVVMFLGLVAVALQLLAFSGWRLTLNL
VILFVLGSGSVVPGFSPARRWVRQVPTHYAVVTIRGLVLRGSPISAYADWLGCL
VAAVGAALVGAILEYERRA"
complement(7138..8229)
/gene="VNG1250H"
complement(7138..8229)
```

```
CDS
complement(7138..8229)
/gene="VNG1250H"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1250h"
/protein_id="AAG19609.1"
/db_xref="GI:10580777"
/translation="MYRDLHVPRTALLAVGLAVLVAAGCTGMRGSSADATTNKL
DSVPASAEYAVADVAGMTDIDLRVYTTTTLAEVULGSRABSDVSSMLDFFEDADI
NPKRIQNTMTFTASTATPTPGTSSGAIVLSTYDESGVNALESSTDTLLNBEYGVN
TVYSDSLGDFDNLVLAALGDGEFVVDGRDSVESMLDVRSGDADHLGDLASFEVNG
GSYVOYALSPNGSDSLIGERYAENAPVNLASVSGIETAYAVSFGTDDGDITSTEHY
ADESGAERAYDIDGTRSLYASPSPEELRPSPEAGSIEQDNTVIVTVAESADTIES
RIEABARTETADNTTTLNSSAASVAP"
complement(8313..11153)
/gene="hela"
/note="hela"
/note="synonym: VNG1251G"
complement(8313..11153)
/gene="hela"
/note="hela"
/codon_start=1
/transl_table=11
/product="ATP-dependent helicase"
/protein_id="AAG19610.1"
/db_xref="GI:10580778"
/translation="MSETGGKMDAFTHLGSEVRNALSRRGTTPTPEPORAVPALD
GRNGLVAPFTGTGKTFLMLPVFDALVSDGFPDGCALVYTPRLRNMRRLDMW
GETLLGEVDVRHGDITTDYQRQDPTDPDVLTPPTLQAMLTGSLRGLADVEHY
VDEHBLAGAKRGAOLAIGLERLVEVAGVQRIGLSATYQDPHEIGAFLTGRGCEII
```

Query Match 18 %; Score 58.2; DB 1; Length 16905;  
Best Local Similarity 53.8%; Pred. No. 27; Mismatches 103; Indels 0; Gaps 0;  
Matches 120; Conservative 0;

QY	28	CTCGCGCGGAGCTCGACGCGCTTCGACGCGGACCGCTCTCGACACGCGTCCGCGCGCATC	87
Db	3865	CTCCGCCCGGAGATCGAGACATCGACCGGAAATCGATCGATCGCGCGACGACG	3806
QY	88	GACCTCGGTGTCGCGATCGCGCGGTACACAGTCCCGGACGCGCTCCGATGATCGAGCC	147
Db	3805	TACGTGCGCGAGACCAATCGCGCGAGTGAAGACCGAGCGGGCTGTGACGACGACGAG	3746
QY	148	GCGCGGTGACCTGTGTCAGACGACGAGCGCCCGCTACGCGCGCCGACCAAGCGCTGAC	207
Db	3745	TCCACGAGGACCCAGTATGATGAGCGCGCCGCGAGAACCCGCGGCTTCGGCGTTCAC	3686
QY	208	GAATCGTTCCTGGTGAACCTCTACGACGTGATCATCACGAGAGA	250
Db	3685	GACCACTCGTGAAGCGGCTTTCGGCTCTCATCGATTGA	3643

RESULT 14  
AP006840\_33/c  
WPCOMMENT

Sequence split into 36 fragments	LOCUS	AP006840	Accession	AP006840
Fragment Name	Begin	End		
AP006840_00	1	110000		
AP006840_01	100001	210000		
AP006840_02	200001	310000		
AP006840_03	300001	410000		
AP006840_04	400001	510000		
AP006840_05	500001	610000		
AP006840_06	600001	710000		
AP006840_07	700001	810000		
AP006840_08	800001	910000		
AP006840_09	900001	1010000		
AP006840_10	1000001	1110000		
AP006840_11	1100001	1210000		
AP006840_12	1200001	1310000		
AP006840_13	1300001	1410000		
AP006840_14	1400001	1510000		
AP006840_15	1500001	1610000		



Query Match	17.9%	Score 55.8	DB 1	Length 110000
Best Local Similarity	55.4%	Pred. No. 36		
Matches 108	Conservative 0	Mismatches 87	Indels 0	Gaps 0
Qy	31	CGCGCGAGCTCGAGCGCCCTCGACCGGACGCTCTCGACACAGTGTGCGCGCGCATCGAC	90	
Db	29205	CGGAGAGGAGATCGACCGCGCTTGGACGAGCGATCTTGGACCTGCTCTGCGCCGATGCG	29146	
Qy	91	CTCGGTGTCGCATCGCGCGGTACAGTCCCGGACGAGCGGTCCGATGATGACGCCGCGC	150	
Db	29145	GCGCGCGCGCCGCGGTGGGCGAGCTGACGCGCGCGCGCGCTGCGCGGTCTGCGAGAGAGAC	29086	
Qy	151	CGGGTCAGCCTGTGTCACAGAGAGGCGCCCGCTCTACGCGCGGACCAACGCGCTTGACGA	210	
Db	29085	CGCAGACACGCGCTCTATCAACCGGCTGCGCGCTGCGCGGAGCGAGCTGCGCGGAG	29026	
Qy	211	TCGTTCTCTGTGTAAC	225	
Db	29025	GAGGTCCGCGGAGATC	29011	
RESULT 15				
SC039120/c				
LOCUS	248550 bp	DNA	linear	BCT 11-FEB-2003
DEFINITION	Streptomyces coelicolor A3(2) complete genome, segment 17/29.			
ACCESSION	AL3939120 AL078618 AL160312 AL160431 AL353815 AL356334 AL392146			
VERSION	AL392146 AL392175 AL392176 AL392178 AL645882			
KEYWORDS	AL3939120.1 GI:24413871			
SOURCE	Streptomyces coelicolor A3(2)			
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.			
REFERENCE	1			
AUTHORS	Bentley, S.D., James, K.F., Cerdano-Tarraga, A.M., Challis, G.L., Thomson, N.R., Bates, K.D., Harris, D.E., Quail, M.A., Kleser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, J., Hladig, J., Hornsby, T., Howarth, S., Huang, C.H., Kleser, T., Larke, J., Murphy, L., Oliver, K., O'Neill, S., Rabinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrett, B.G., Parkhill, J., and Hopwood, D.A.			
TITLE	Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)			
JOURNAL	Nature 417 (6885), 141-147 (2002)			
MEDLINE	21996410			
PUBMED	12000953			
REFERENCE	2 (bases 1 to 248550)			
AUTHORS	Bentley, S.D.			
TITLE	Direct Submission			

JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA E-mail: sds@sanger.ac.uk

COMMENT On or before Oct 26, 2002 this sequence version replaced gi:11595527, gi:7242710, gi:17245322, gi:20520808, gi:7581386, gi:10129696, gi:10129727, gi:10178320, gi:10178340, gi:10178373.

FEATURES  
Source Location=Qualifiers  
1..248550 /organism="Streptomyces coelicolor A3(2)"  
/mol\_type="genomic DNA"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
406..897 /gene="SCO4412"  
/note="synonym: SCGF11.10"  
406..897 /gene="SCO4412"  
/note="synonym: SCGF11.10"  
406..897 /gene="SCO4412"  
/note="SCGF11.10, possible regulatory protein, len: 163 aa, similar to TR\_Q9ZBR2 (EMBL:AI034446) Streptomyc coelicolor putative regulatory protein SCIA9.20, 135 aa; fasta scores: opt: 227 z-score: 286.6 E(): 1.9e-08; 43.0% identity in 135 aa overlap"  
/codon\_start=1  
/transl\_table=11  
/product="putative regulatory protein"  
/protein\_id="CAC08421.1"  
/db\_xref="GI:10178383"  
/db\_xref="GOA:G9F376"  
/db\_xref="UniProt/TRMBL\_09P376"  
/translation="MOVLYQLEVRPDPESEVGARKMAWSRLAGSSIRADEPLAEVLVLSSELTAIVAVHTRGSVAIRLVPGAVIEETDEATVRLVEDDASRADVPKCAEGSTGRGALVDGLADRGWMSREGAKRIWCLELDRAAGEAPVCGSGAVAAEGFAYEVN"  
complement (1200..2234)  
/gene="SCO4413"  
/note="synonym: SCGF11.11c"  
complement (1200..2234)  
/gene="SCO4413"  
/note="SCGF11.11c, probable arac family transcriptional regulatory protein, len: 344 aa; highly similar to TR\_Q9SI66 (EMBL:AB023785) Streptomyces griseus AdpA, 405 aa; fasta scores: opt: 881 z-score: 1024.1 E(): 0; 46.9% identity in 322 aa overlap and to TR:CAB70646 (EMBL:AL137242) Streptomyces coelicolor putative transcriptional regulator SCGF4.20c, 337 aa; fasta scores: opt: 1237 z-score: 1238.7 E(): 0; 60.3% identity in 325 aa overlap. Contains pfam match to entry PF00165 HTH\_ARAC, Bacterial regulatory helix-turn-helix proteins, arac family and a possible helix-turn-helix motif at residues 249..270 (+2.55 SD)"  
/codon\_start=1  
/transl\_table=11  
/product="putative arac family transcriptional regulatory protein"  
/protein\_id="CAC08422.1"  
/db\_xref="GI:10178384"  
/db\_xref="GOA:G9F375"  
/db\_xref="UniProt/TRMBL\_09P375"  
/translation="MAPETLSAPGAHGAVRPHRVVTLALDDLPPELGISAIRFELA RDLHRPLLEVTVCSVRPPGPAVTDADEFILVNGGEALATADTVPVPSYELGPVKE EGTLTGELSAALAHLAVSPGRILVAICTGGCVFLAAGYLDERPATTHWASEHFQRVYK AVTDADVLLFVDDGDVLTAGVAAGIDLCGHIVLRDRHGAVANVARRTVAPRRDGGO AXVIERPVDPPOOSTTARVALRGILRHPIOLDMAAQESMVSPTFRPRFEVGVSGWMLTRQRYERRRHLLESTALSVDVOADDAGGTGQSNRQLQAVLGATPTGYRRTF RAFLSTVACE"  
complement (1233..1493)  
/gene="SCO4413"  
/note="pfam match to entry PF00165 HTH\_ARAC, Bacterial regulatory helix-turn-helix proteins, arac family, score 74.90, E-value 1.6e-18"  
2305..2308  
2319..3644 /gene="SCO4414"



```

CDS
/note="synonym: SC6F11.12"
/2319..3644
/gene="SCO4414"
/feature="SC6F11.12, possible integral membrane protein, len:
441 aa; similar to TR:BA06413 (EMBL:AP001516) Bacillus
halodurans transporter BH2694, 418 aa; fasta scores: opt:
795 z-score: 909.8 E(): 0; 31.0% identity in 413 aa
overlap. Contains possible hydrophobic membrane spanning
regions"
/codon_start=1
/translation="putative integral membrane transport protein"
/protein_id="CAC08423.1"
/db_xref="GI:10178385"
/feature="UniProt/TrEMBL:Q9F374"
/translation="MTOTDAAAAAASERPEPTRRGPTRAHRAWVAATPVITII
GAAPRSLPGILIDPLHOFQFMSRGTGAASVVALYGLTAPPAALMDRFGIRRV
AVALLMVALSGSLTVNMSAMQILICGLVGLSGSMALFAATVTNMTFRGLV
TGVLTASASGQILFLPLSMIVETVDMRPAATVALAALAVPFWLLDHPADV
QKPYGAKEFVKKPAPVPGARBALTVFAVRGPFMLAGTALICGASTNGLIOTHF
VPAADHGMPTTAASLILAVICFVDVAGTASCFMTRPERRLIAYVALRVSILF
LPMULAPVHPMIFLTVFYGLDVAIVPTLALCRHVEDSAIVFGWVLASHQVCA
ALVAFLEGVADRFGSYDVVVIASGALCAAAALMALVIRRRPVVALS"
/3666..4538
/gene="SCO4415"
/feature="synonym: SC6F11.13"
/3666..4538
/gene="SCO4415"
/feature="SC6F11.13, possible integral membrane protein, len:
290 aa. Contains possible hydrophobic membrane spanning
regions"
/codon_start=1
/translation="putative integral membrane protein"
/protein_id="CAC08424.1"
/db_xref="GI:10178386"
/feature="UniProt/TrEMBL:Q9F373"
/translation="MGWGTPEEAEGRPMSNDENVAALCATQPLMFLMWPRV
GAQDDHGRGGAGFVLCPLVPLGLSVSSAQIMPAATLALIPRSSEPARHV
ASVAVGAMVAVLHVLWMPRPDTPVFAAGILPVFLPLRRRAGMSVLRSG
ACFVLVAGGLAAVALADEYEPALSDGQLAGEMRGHGVLLAPGGRGLTRVSR
PCEGTADYTRCGGTGTWSRNLSDRGVAVRLDGCGETSWTIGSSERSPLFLAF
GDDPAGDLFLTN"
/complement(4721..5926)
/gene="SCO4416"
/feature="synonym: SC6F11.14c"
/complement(4721..5926)
/gene="SCO4416"
/feature="SC6F11.14c, probable monooxygenase, len: 401 aa;
similar to TR:Q95D6 (EMBL:AP000837) Oryza sativa similar
to human dimethylallyl pyrophosphate monooxygenase, 437 aa; fasta
scores: opt: 692 z-score: 722.8 E(): 9.3e-33; 38.2%
identity in 403 aa overlap. Contains Pfam match to entry
PF00743 FMO-like, Flavin-binding monooxygenase-like"
/codon_start=1
/translation="putative monooxygenase"
/protein_id="CAC08425.1"
/db_xref="GI:10178387"
/db_xref="GOA:Q9F372"
/feature="UniProt/TrEMBL:Q9F372"
/translation="MADSTASAPQRPPTHAPRPVYVIGAPGLAVAAHARGLRA
VLERADHVGSSWRHRYDLRLHTRRLSALPGIPRRGRVAVDDVRYLEKTAE
YHOLEIVTGVFVERAPDGTGMLLHAAGRLTGAVVATGVNHTPRVPMGPD
TYTGEPRHADYRTAPYAGDVAVGVNTGABIVDIVEGGAARVLAVRTAPHHV
RSTAGAAOYTGVCRLPVLPALVDPLARPLAISVDLSAOGCLPRGTGLSRVAVG
AIPVQVGLIDAVRSGRVAVVAMDGFEDEKVLADSTRAPDAVIAATGIRKLESL
AGVLSRLPG"
/complement(4754..5875)
/gene="SCO4416"
/feature="Pfam match to entry PF00743 FMO-like,
flavin-binding monooxygenase-like, score -255.00, E-value
```

```

gene
8..8e-09"
6026..6643
/feature="SCO4417"
/feature="synonym: SC6F11.15"
6026..6643
CDS
/feature="SCO4417"
/feature="SC6F11.15, probable acetyltransferase, len: 205 aa;
similar to SW:RIWL_ECOLI (EMBL:X15860) Escherichia coli
ribosomal-protein-terine acetyltransferase (EC 2.3.1.-)
RimU, 179 aa; fasta scores: opt: 199 z-score: 261.1 E():
4.8e-07; 30.6% identity in 157 aa overlap. Contains Pfam
match to entry PF00583 Acetyltransf, Acetyltransferase
(GNAT) family"
/codon_start=1
/translation="putative acetyltransferase"
/protein_id="CAC08426.1"
/db_xref="GI:10178388"
/db_xref="GOA:Q9F371"
/feature="UniProt/TrEMBL:Q9F371"
/translation="MAHTPMYTLDESLSRTERRGHGLRLPAMDAGSDDAHTVRLCL
DPEFQWNTTLKMLTDLGARSLSLRADADGRASFRITDESGTTLGHILNEI
SRQMKYARGVYVLPPEARGRGVALRALVASSWAVYGLHRLLELGHAVGHASCRVA
ERSGVYVEGTLREAMEAGRHDAFRVHLHARLATDEBPTPGAR"
/6290..6523
/gene="SCO4417"

misc_feature
Query Match 17.9% Score 55.8 DB 1; Length 248550;
Best Local Similarity 49.8%; Pred. No. 30;
Matches 141; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 9 GCAGAACGAGCTGACAGCGGTGGCGGAGACTCGACGCCCTTGACGAGGACTCCTGCA 68
Db 185684 GCTGTTCTGGCCCTCTCAAGCCCGGAGCGGAGAGTCTCCGCCACCGAGTGGT 185625

QY 69 CACGATGCGGCGCGCGATGACCTGATGCGGATGCGGATGCGGATGCGGATGCGG 128
Db 185624 GTGATCCTCGCTTGTGACCTGCGCTCTCTTGTGACGCGCTGGGCTGGGCGG 185565

QY 129 CGTCCGATGATGACAGCCCGCGGCGGTGACCTGATGACGACGAGCGCGCGCTTACGC 188
Db 185564 CGAGCTGCTGCGGCGGACCGCGGCGGTGACCTGATGACGAGCGCGCGCTTACGC 185505

QY 189 CGCCGACGAGCGCTTGAAGATGCTTCTGTTGAACTTTCAGAGCTATCAGCA 248
Db 185504 CGTCACTGGGCGGTCTATCATCTGCGCGCTGAAACAGCGCGCATGCTGAGCGCTTGCCT 185445

QY 249 GATGTGCGCGGTGAGGAGCTGTGATGAGCGGAGGAGCGCTG 291
Db 185444 CGGCTACTTATCAACCGCGCTGTCACATCGGATGGCGGTG 185402

Search completed: October 5, 2005, 05:30:50
Job time : 1437.22 secs
```



This Page Blank (uspio)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 16:46:09 : Search time 188.488 Seconds  
(without alignments)  
9798.799 Million cell updates/sec

Title: US-10-089-514-3

Perfect score: 312  
Sequence: 1 atgcacgcagacagacagct.....cggccgagacgcgcggtga 312

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: N\_Geneseq\_16Dec04:\*  
2: geneseqn1980s:\*  
3: geneseqn1990s:\*  
4: geneseqn2000s:\*  
5: geneseqn2001as:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	312	100.0	312	4	AAF86410
2	312	100.0	312	8	AA150181
3	296.4	100.0	3305	8	ABZ68799
4	264.8	84.9	12391	8	ABZ68798
5	135.6	43.5	387	2	AAT58552
6	135.6	43.5	2888	2	AAT59268
7	54.8	17.6	1458	10	ABZ66707
8	54.8	17.6	45055	10	ABZ66808
9	54.6	17.5	507	11	ABD12889
10	54.6	17.5	1359	11	ABD12819
11	54.6	17.5	1371	11	ABD12866
12	54.6	17.5	73882	13	AD873531
13	54	17.3	1208	2	AAQ64201
14	54	17.3	5392	2	AAQ64201
15	53.6	17.2	837	10	ABZ66725
16	53.6	17.2	1933	2	AAT43282
17	53.6	17.2	29870	8	AAD36874
18	53.6	17.2	45055	10	ABZ66808
19	53	17.0	1356	8	ACA42290
20	52.8	16.9	510	11	ABD09571

21	52.8	16.9	1116	11	ABD09634	Abd09634 Pseudomon
22	52.8	16.9	1125	11	ABD09470	Abd09470 Pseudomon
23	52.6	16.9	1869	13	AD55461	Ad55461 Bacterial
24	52.6	16.9	1869	13	AD549946	Ad549946 Bacterial
25	52.6	16.9	3735	13	ADT45696	Adt45696 Bacterial
26	52.4	16.8	1410	12	ADM80073	Adm80073 Spiroamyl
27	52.4	16.8	1410	12	ADN97589	Adn97589 S ambofac
28	52.4	16.8	30943	12	ADM80034	Adm80034 Spiroamyl
29	52.4	16.8	30943	12	ADN97550	Adn97550 S ambofac
30	51.8	16.6	1686	8	AD36877	Ad36877 S. clavul
31	51.6	16.5	13654	10	ADC86678	Adc86678 Human GPC
32	51.4	16.5	388	9	ACL18903	ACL18903 DNA clone
33	51.2	16.4	313	4	AAC89250	Aac89250 Human bra
34	51.2	16.4	5469	2	AAV83491	Aav83491 Human T-t
35	51.2	16.4	5562	2	AAV57542	Aav57542 Human cal
36	51.2	16.4	5562	5	AAF31674	Aaf31674 Human alp
37	51.2	16.4	6816	6	AA516826	Aa516826 Human T-t
38	51.2	16.4	6816	8	ABX93560	Abx93560 Human CDN
39	51.2	16.4	6816	12	ADH69264	Adh69264 Human TTC
40	51.2	16.4	6855	6	AA516827	Aa516827 Human T-t
41	51.2	16.4	6855	8	ABX93561	Abx93561 Human CDN
42	51.2	16.4	6855	12	ADH69266	Adh69266 Human TTC
43	51.2	16.4	6990	13	AD516296	Ad516296 Human vol
44	51.2	16.4	25467	12	ADQ07566	Adq07566 Streptomy
45	51	16.3	113193	8	AAD54645	Aad54645 Streptomy

#### ALIGNMENTS

##### RESULT 1

AAF86410  
ID AAF86410 standard; DNA; 312 BP.

AC AAF86410;

DT 22-JUN-2001 (first entry)

DE 4-amino-4-deoxychorismic acid mutase coding sequence.

XX Metabolite; benzene; chorismic acid; p-aminophenylpyruvic acid;

KM 4-amino-4-deoxychorismic acid mutase; enzyme; papB; ds.

XX Streptomyces venezuelae.

OS Key Location/Qualifiers

FT CDS 1..312

FT /product= "4-amino-4-deoxychorismic acid mutase"

XX WO200123542-A1.

XX 05-APR-2001.

PF 29-SEP-2000; 2000WO-JP006783.

XX 29-SEP-1999; 99JP-00276314.

XX	(MEIU ) MEIJI SEIKA KAISHA LTD.
PA	Yanaï K, Okakura K, Yaeuda S, Watanabe M, Miyamoto K, Micho N;
PI	Murakami T;
DR	WPI; 2001-290517/30.
DR	P-PSDB; AAB82072.
XX	Transformant producing secondary metabolite modified with functional
PT	group e.g. benzene with nitrogen-containing substituent at para-position,
PT	PF1022, with ease at low cost, for application in pharmaceuticals and
XX	agrochemicals.
XX	Claim 13; Page 65; 83pp; Japanese.



CC The present invention relates to a transformant having been modified so  
 CC as to produce a secondary metabolite. The secondary metabolite has a  
 CC benzene ring skeleton free from substitution at the para-position by a  
 CC nitrogen-containing functional group, thereby enabling the production of  
 CC a secondary metabolite with a benzene ring skeleton substituted at the  
 CC para-position by a nitrogen-containing group. The transformant organism  
 CC participating in the biosynthesis pathway from chorismic acid into p-  
 CC aminophenylpyruvic acid. The present sequence is the coding sequence for  
 CC 4-amino-4-deoxychorismic acid mutase (pabp), from Streptomyces  
 CC venezuelae. pabp participates in the biosynthesis pathway from chorismic  
 CC acid into p-aminophenylpyruvic acid, and so the pabp gene can be used to  
 CC produce the transformant of the present invention. The transformant can  
 CC be used to produce metabolites for application in pharmaceuticals,  
 CC veterinary drugs and agrochemicals  
 XX  
 SQ Sequence 312 BP; 54 A; 106 C; 111 G; 41 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 312; DB 4; Length 312;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-46;  
 Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGACCGAGCAGAACGAGCTGACAGCGGCTGCGCGGAGCTCGACCGCTTCGACGGGACG 60  
 DB 1 ATGACCGAGCAGAACGAGCTGACAGCGGCTGCGCGGAGCTCGACCGCTTCGACGGGACG 60  
 QY 61 CTCTTGACACAGGATGCGCGCGCCGATGACCTCGATGTCGCACTCGCGGATCAAGTCC 120  
 DB 61 CTCTTGACACAGGATGCGCGCGCCGATGACCTCGATGTCGCACTCGCGGATCAAGTCC 120  
 QY 121 CGGACGCGGCTCCCGATGATGACGCGCGGCTGAGCTGATCAAGACAGGCGCGCC 180  
 DB 121 CGGACGCGGCTCCCGATGATGACGCGCGGCTGAGCTGATCAAGACAGGCGCGCC 180  
 QY 181 CGCTAGCGCGCGCCGACCAAGCGGCTCGAGGATCGTCTCGGTGAACCTTACGACGATGC 240  
 DB 181 CGCTAGCGCGCGCCGACCAAGCGGCTCGAGGATCGTCTCGGTGAACCTTACGACGATGC 240  
 QY 241 ATCAGCGAGATGTGCGCGCTCGAGGACCTGTGATGAGCGGAGAGAGCTTACGCGCGAG 300  
 DB 241 ATCAGCGAGATGTGCGCGCTCGAGGACCTGTGATGAGCGGAGAGAGCTTACGCGCGAG 300  
 QY 301 GACCGGCGGTGA 312  
 DB 301 GACCGGCGGTGA 312  
 RESULT 2  
 AAL50181  
 ID AAL50181 standard; DNA; 312 BP.  
 AC AAL50181;  
 XX  
 DT 28-JAN-2003 (first entry)  
 XX  
 DE S venezuelae pf1022 substance gene #2.  
 XX  
 KM Substance pf1022; biosynthesis; transgenic; chorismic acid; anthelmintic;  
 KM phenylalanine; p-aminophenylpyruvic acid; pharmaceutical;  
 KM veterinary drug; gene; ds.  
 XX  
 OS Streptomyces venezuelae.  
 XX  
 FT Key Location/Qualifiers  
 FT CDS 1..312  
 FT /\*tag= a  
 FT /product= "pf1022 substance"  
 XX  
 PN MO200277244-A1.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 22-MAR-2002; 2002WO-JP002782.

XX  
 PR 22-MAR-2001; 2001JP-00082227.  
 XX  
 PA (MEIJ ) MEIJI SEIKA KAISHA LTD.  
 XX  
 PT Yanai K, Sumida N, Watanabe M, Moriya T, Murakami T;  
 XX  
 DR WPI; 2003-018934/01.  
 DR P-PSDB; AAO19564.  
 PT Novel biosynthesis gene-transferred transformants for producing pf1022  
 PT substance derivatives by fermentation, as pharmaceuticals or veterinary  
 PT drugs with anthelmintic activity.  
 XX  
 PS Claim 15; Page 85; 116pp; Japanese.  
 XX  
 CC The present invention relates to transformants capable of producing  
 CC pf1022 substance derivatives. These were obtained by transferring a gene  
 CC participating in the biosynthesis pathway from chorismic acid to p-  
 CC aminophenylpyruvic acid (biosynthesis gene) into a phenylalanine-  
 CC requiring host derived from an organism producing the pf1022 substance.  
 CC The transformants are producing pf1022 substance derivatives by  
 CC fermentation, for use as pharmaceuticals or veterinary drugs. The present  
 CC sequence is a substance pf1022 coding sequence from Streptomyces  
 CC venezuelae  
 XX  
 SQ Sequence 312 BP; 54 A; 106 C; 111 G; 41 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 312; DB 8; Length 312;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-46;  
 Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGACCGAGCAGAACGAGCTGACAGCGGCTGCGCGGAGCTCGACCGCTTCGACGGGACG 60  
 DB 1 ATGACCGAGCAGAACGAGCTGACAGCGGCTGCGCGGAGCTCGACCGCTTCGACGGGACG 60  
 QY 61 CTCTTGACACAGGATGCGCGCGCCGATGACCTCGATGTCGCACTCGCGGATCAAGTCC 120  
 DB 61 CTCTTGACACAGGATGCGCGCGCCGATGACCTCGATGTCGCACTCGCGGATCAAGTCC 120  
 QY 121 CGGACGCGGCTCCCGATGATGACGCGCGGCTGAGCTGATCAAGACAGGCGCGCC 180  
 DB 121 CGGACGCGGCTCCCGATGATGACGCGCGGCTGAGCTGATCAAGACAGGCGCGCC 180  
 QY 181 CGCTAGCGCGCGCCGACCAAGCGGCTCGAGGATCGTCTCGGTGAACCTTACGACGATGC 240  
 DB 181 CGCTAGCGCGCGCCGACCAAGCGGCTCGAGGATCGTCTCGGTGAACCTTACGACGATGC 240  
 QY 241 ATCAGCGAGATGTGCGCGCTCGAGGACCTGTGATGAGCGGAGAGAGCTTACGCGCGAG 300  
 DB 241 ATCAGCGAGATGTGCGCGCTCGAGGACCTGTGATGAGCGGAGAGAGCTTACGCGCGAG 300  
 QY 301 GACCGGCGGTGA 312  
 DB 301 GACCGGCGGTGA 312  
 RESULT 3  
 AB269799  
 ID AB269799 standard; DNA; 3305 BP.  
 AC AB269799;  
 XX  
 DT 08-APR-2003 (first entry)  
 XX  
 DE Plasmid pABABC.  
 XX  
 KM DHFR; translation; orthogonal tRNA; O-tRNA; O-RS; CAT; TYRS;  
 KM orthogonal aminoacyl tRNA synthetase; unnatural amino acid;  
 KM chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;  
 KM biosynthesis; p-aminophenylalanine; pAF; chorismate.  
 XX  
 OS Synthetic.



```
XX MO200285923-A2.
PN
XX
XX 31-OCT-2002.
PD
XX 19-APR-2002; 2002MO-US012465.
PF
XX 19-APR-2001; 2001US-0285030P.
PR 06-FEB-2002; 2002US-0355514P.
XX
PA (SCRI ) SCRIPPS RES INST.
PI Schultz P, Wang L, Anderson JC, Chin JWK, Liu DR, Magliery TJ;
PI Meggers EL, Mehl RA, Pastinak M, Santoro SW, Zhang Z;
XX WPI; 2003-120430/11.
DR
XX
XX Composition useful for producing protein comprising unnatural amino acid,
PT has translation system comprising orthogonal tRNA and orthogonal
PT aminoacyl tRNA synthetase.
XX
XX Example 4; Page 127-128; 188pp; English.
XX
XX The invention relates to a novel composition comprising a translation
CC system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl
CC tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O
CC -tRNA with at least one unnatural amino acid in the translation system
CC and the O-tRNA recognises at least one selector codon. A composition of
CC the invention is useful for producing at least one protein comprising at
CC least one unnatural amino acid. The protein is the Asp127AG mutant of
CC chloramphenicol acetyltransferase (CAT), the Tyr163TAG mutant of mouse
CC dihydrofolate reductase (DHFR), or a Tyr163TAG mutant of mouse
CC dihydrofolate reductase comprising CHO0H His6tag. The unnatural amino
CC acid is provided exogenously. The translation system is a cell and the
CC unnatural amino acid is biosynthesised by the cell. The present sequence
CC represents a plasmid containing the individual genes papABC that encode
CC the enzymes used to carry out the conversion of chorismate to the
CC unnatural amino acid p-aminophenylalanine (pAF)
XX
SQ Sequence 3305 BP; 445 A; 1327 C; 1121 G; 412 T; 0 U; 0 Other;
Query Match 95.0%; Score 296.4; DB 8; Length 3305;
Best Local Similarity 98.4%; Pred. No. 6.6e-44;
Matches 311; Conservative 0; Mismatches 1; Indels 4; Gaps 1;
QY 1 ATGACCGAGCAAGAACGAGCTGCA---GCGGCTGCGCGGAGCTCGACGCCCTCGACGG 56
DB 2990 ATGACCGAGCAAGAACGAGCTGCGGCTGCGCGGAGCTCGACGCCCTCGACGG 3049
QY 57 GAGGCTCTGGAACACGGTGGCGGCGCGCATCGACTCGGTGTCGGCATCGCGCGGTACAA 116
DB 3050 GAGGCTCTGGAACACGGTGGCGGCGCGCATCGACTCGGTGTCGGCATCGCGCGGTACAA 3109
QY 117 GTCCCGGACACGGGCTCCCGATGATGACAGCCCGGCGGGTCACTGATCAAGACAGGGC 176
DB 3110 GTCCCGGACACGGGCTCCCGATGATGACAGCCCGGCGGGTCACTGATCAAGACAGGGC 3169
QY 177 CGCCCGCTACGCGCGCGACCAAGGCTTCGAGCAATGTTCTGTGTAACCTTACAGAGT 236
DB 3170 CGCCCGCTACGCGCGCGACCAAGGCTTCGAGCAATGTTCTGTGTAACCTTACAGAGT 3229
QY 237 GATCATACAGGAGATGTCGCGCGTGAAGACCTGTGATGACCGGAGAGAGCTCGACGGC 296
DB 3230 GATCATACAGGAGATGTCGCGCGTGAAGACCTGTGATGACCGGAGAGAGCTCGACGGC 3289
QY 297 CGAGAACCGGCGGTGA 312
DB 3290 CGAGAACCGGCGGTGA 3305
RESULT 4
AB269798
ID AB269798 standard; DNA; 12391 BP.
```

```
XX AB269798;
AC
XX
XX 08-APR-2003 (first entry)
DT
XX
XX Plasmid plasc-papabc.
DE
XX
XX DHFR; translation; orthogonal tRNA; O-tRNA; O-RS; CAT; TyrRS;
KW orthogonal aminoacyl tRNA synthetase; unnatural amino acid;
KW chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;
KW biosynthesis; p-aminophenylalanine; pAF.
XX
XX Synthetic.
OS
XX
XX MO200285923-A2.
PN
XX
XX 31-OCT-2002.
PD
XX 19-APR-2002; 2002MO-US012465.
PF
XX 19-APR-2001; 2001US-0285030P.
PR 06-FEB-2002; 2002US-0355514P.
XX
XX (SCRI ) SCRIPPS RES INST.
XX
XX Schultz P, Wang L, Anderson JC, Chin JWK, Liu DR, Magliery TJ;
PI Meggers EL, Mehl RA, Pastinak M, Santoro SW, Zhang Z;
XX WPI; 2003-120430/11.
DR
XX
XX Composition useful for producing protein comprising unnatural amino acid,
PT has translation system comprising orthogonal tRNA and orthogonal
PT aminoacyl tRNA synthetase.
XX
XX Example 4; Page 124-127; 188pp; English:
XX
XX The invention relates to a novel composition comprising a translation
CC system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl
CC tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O
CC -tRNA with at least one unnatural amino acid in the translation system
CC and the O-tRNA recognises at least one selector codon. A composition of
CC the invention is useful for producing at least one protein comprising at
CC least one unnatural amino acid. The protein is the Asp127AG mutant of
CC chloramphenicol acetyltransferase (CAT), the Tyr163TAG mutant of mouse
CC dihydrofolate reductase (DHFR), or a Tyr163TAG mutant of mouse
CC dihydrofolate reductase comprising CHO0H His6tag. The unnatural amino
CC acid is provided exogenously. The translation system is a cell and the
CC unnatural amino acid is biosynthesised by the cell. The present sequence
CC represents a plasmid for use in the biosynthesis of p-aminophenylalanine
CC (pAF) in vivo
XX
SQ Sequence 12391 BP; 2830 A; 3588 C; 3346 G; 2627 T; 0 U; 0 Other;
Query Match 84.9%; Score 264.8; DB 8; Length 12391;
Best Local Similarity 97.9%; Pred. No. 2.3e-38;
Matches 280; Conservative 0; Mismatches 2; Indels 4; Gaps 1;
QY 1 ATGACCGAGCAAGAACGAGCTGCA---GCGGCTGCGCGGAGCTCGACGCCCTCGACGG 56
DB 2055 ATGACCGAGCAAGAACGAGCTGCGGCTGCGCGGAGCTCGACGCCCTCGACGG 2114
QY 57 GAGGCTCTGGAACACGGTGGCGGCGCGCATCGACTCGGTGTCGGCATCGCGCGGTACAA 116
DB 2115 GAGGCTCTGGAACACGGTGGCGGCGCGCATCGACTCGGTGTCGGCATCGCGCGGTACAA 2174
QY 117 GTCCCGGACACGGGCTCCCGATGATGACAGCCCGGCGGGTCACTGATCAAGACAGGGC 176
DB 2175 GTCCCGGACACGGGCTCCCGATGATGACAGCCCGGCGGGTCACTGATCAAGACAGGGC 2234
QY 177 CGCCCGCTACGCGCGCGACCAAGGCTTCGAGCAATGTTCTGTGTAACCTTACAGAGT 236
DB 2235 CGCCCGCTACGCGCGCGACCAAGGCTTCGAGCAATGTTCTGTGTAACCTTACAGAGT 2294
```



QY	227	GATCATCAACGAGATGTCCGGCTGCAGACA	CCTGTGGAGACCGG	282
Dd	2295	GATCATCAACGAGATGTCCGGCTGCAGACA	CCTGTGGAGACCGG	2240
 RESULT 5				
ID	AAT58552	AAT58552 standard; cDNA, 387 BP.		
AC	AAT58552;			
DT	02-APR-1997	(first entry)		
DE	Streptomyces pristinaespiralis papB gene.			
XX				
KW	Streptogramin B; antibiotic; biosynthesis; pristinamycin; virginiamycin;			
KV	DMPAPA precursor; 4-dimethylamino-L-phenylalanine; papA; papM; papB;			
KW	papC; isomerisation; aromatisation; N-methyltransferase; ds.			
XX				
OS	Streptomyces pristinaespiralis.			
FX				
FH	Key	Location/Qualifiers		
FT	mat_peptide	1..387		
FT		/*tag = a		
FT		/product= "Paps"		
XX				
PN	WO9601901-A1.			
PD	25-JAN-1996.			
XX				
PF	04-JUL-1995;	95MO-FR000889.		
PR	08-JUL-1994;	94FR-00008478.		
PA	(RHON ) RHONE POULENC RORER SA.			
PI	Blanc V, Thibaut D, Bamas-Jacques N, Blanche F, Crouzet J,			
PL	Barriere J, Debussche L, Famechon A, Paris J, Dutruc-Rosset G,			
DR	WPI; 1996-097631/10.			
XX	P-PSDB; AAM11583.			
PT	New streptogramin B derivs. useful as antibiotics - produced by new			
PT	mutants of Streptomyces having altered genes for streptogramin B			
XX	biosynthesis.			
PS	Claim 18; Page 106-107; 146pp; French.			
XX				
CC	The papA and papM genes of S.pristinaespiralis are involved in the			
CC	biosynthesis of 4-dimethylamino-L-phenylalanine (DMPAPA), a precursor for			
CC	pristinamycin IA. The region between these two genes was sequenced and			
CC	two open reading frames were identified. The first (papC) was on the			
CC	complementary strand and decodes to an amino acid sequence with homology			
CC	to the region of B.coli TyrA which has been implicated in aromatisation			
CC	reactions. The papC gene product is likely to be involved in a similar			
CC	aromatisation of 4-deoxy 4-amino prephenate to give 4-amino			
CC	phenylglyuvate during DMPAPA synthesis. The second open reading frame			
CC	(papB) could be decoded to give a product with homology to the region of			
CC	TyrA which has chorismate mutase activity. The papB gene product is			
CC	likely to catalyse the isomerisation of 4-deoxy 4-amino chorismate to 4-			
CC	deoxy 4-amino prephenate acid. Disruption of the papB and papC genes can			
CC	be used to produce strains of S.pristinaespiralis which are unable to			
CC	produce the antibiotic pristinamycin I but which may be able to produce			
CC	new, modified forms of it			
XX				
SO	Sequence 387 BP; 50 A; 187 C; 110 G; 40 T; 0 U; 0 Other;			
	Query Match	43.5%; Score 135.6; DB 2; Length 387;		
	Best Local Similarity	69.8%; Pred. No. 2,3e-15;		
	Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0			
QY	14	ACGAGCTGCAGCGGCTGCCGGCGGAGACTGCACGCCCTTCAGCGGAGACGCTCTTGGAACACGG	73	

Dd		59	ACCCCTCGACAGCGGCTGGCGGCCGCCCTTGGACGCCGGGGAACGCCGCCCTTGCTGTGGACGCC	118
Oy		74	TGCAGCGCCGCATCGACTCGGTGTGTCGCATTCGCGCGGTACAGATCCCGCAGCGCGTCC	133
Dd		119	TCCGCACAACGCTTGAGCATCTGCTCGGCCATTCGGAGTAGTAACAACGCGCTCCACCAGAGTGC	178
Oy		134	CATTATGTCAGGCCCGCGCGGATCACTCTGTTTCAGAGACAGGGCGCCCCGCTACGCCGCCG	193
Dd		179	CGATGATGAGGCCCCCACCGGATCGCCAGAGTCCACAGCCAAAGCCGCCGCTACGCCGCCG	238
Oy		194	ACCAGCGCCCTCGACGAAATGGTTCCTGTGTAACCTTCAACGATGATCATCAGGAGATGT	253
Dd		239	ACCAACGAGCATCGAACCCCGCTTCTCTGCGCACCTGTACACAGATCATCAGAGACT	298
Oy		254	GCCGCGTCGAGGACCTGTGTAT	275
Dd		299	GCCGCTCGAGGACGAGTGGAT	320
RESULT 6				
AAT59268				
ID		AAT59268	standard; cDNA; 2888 BP.	
XX				
XX		AAT59268;		
XX				
DT		02-APR-1997	(first entry)	
XX				
DE		Streptomyces pristinaespiralis papa and papM intergenic region.		
XX				
KW		Streptogramin B; antibiotic; biosynthesis; pristinamycin; virginiamycin;		
KM		DMAPA precursor; 4-dimethylamino-L-phenylalanine; papa; papM; pabB;		
OS		pabC; isomerisation; aromatization; N-methyltransferase; ds.		
XX				
OS		Streptomyces pristinaespiralis.		
XX				
FH			Location/Qualifiers	
FT			1..687	
FT			/tag= a	
FT			/product= "Papa"	
FT			/note= "C-terminal coding region only, i.e. a partial	
FT			open reading frame"	
FT			complement(949..1836)	
CDS			/tag= b	
FT			/product= "PapC"	
FT			1873..2262	
FT			/tag= C	
CDS			/product= "pabB"	
FT			2259..2888	
FT			/tag= d	
FT			/product= "PapM"	
FT			/note= "N-terminal coding region only, i.e. a partial	
FT			open reading frame"	
PM		WO9601901-A1.		
PD		25-JAN-1996.		
XX				
PE		04-JUL-1995;	95WO-FR000889.	
XX				
PR		08-JUL-1994;	94FR-00088478.	
XX				
PA		(RHON ) RHONE POULENC RORER SA.		
PI		Blanc V, Thibaut D, Bamas-Jacques N,	Blanche F, Crouzet J;	
PI		Barriere J, Debussche L, Famechon A,	Paris J, Dutruc-Rosset G;	
DR		WPi; 1996-097631/10.		
DR		P-PSDB; AAM11582, AAM11583.		
XX		New streptogramin B derivs. useful as antibiotics - produced by new		
PT		mutants of Streptomyces having altered genes for streptogramin B		
PT		biosynthesis.		
XX				



PS Example 1; Page 102-104; 146pp; French.  
XX  
CC The papA and papM genes of *S. pristinaespiralis* are involved in the  
CC biosynthesis of 4-dimethylamino-L-phenylalanine (DMPPA), a precursor for  
CC pristinamycin I<sub>A</sub>. The region between these two genes was sequenced and  
CC two open reading frames were identified. The first (papC) was on the  
CC complementary strand and decodes to an amino acid sequence with homology  
CC to the region of *E. coli* TyrA which has been implicated in aromatization  
CC reactions. The papC gene product is likely to be involved in a similar  
CC aromatization of 4-deoxy 4-amino prephenate to give 4-amino  
CC phenylpyruvate during DMPPA synthesis. The second open reading frame  
CC (papB) could be decoded to give a product with homology to the region of  
CC TyrA which has chorismate mutase activity. The papB gene product is  
CC likely to catalyse the isomerisation of 4-deoxy 4-amino chorismate to 4-  
CC deoxy 4-aminoprephenate. Disruption of the papB and papC genes can be  
CC used to produce strains of *S. pristinaespiralis* which are unable to  
CC produce the antibiotic pristinamycin I but which may be able to produce  
CC new, modified forms of it  
XX  
SQ Sequence 2888 BP; 390 A; 1319 C; 897 G; 282 T; 0 U; 0 Other;  
Query Match 43.5%; Score 135.6; DB 2; Length 2888;  
Best Local Similarity 69.8%; Pred. No. 1.9e-15;  
Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0;  
QY 14 ACGAGCTGCAGCGCGCGCGGAGCTCCAGCGCCCTCCAGCGGACGCTCTCTGACACGG 73  
DB 1931 ACCCCCTCGACGCGCTCGCGCGCGCGCTCGACGCGCGGAGCGCCCTCTGTGACGCG 1990  
QY 74 TGGCGCGCGCGCATCGACTCGGTGTCGCGCATCGCGCGGTACAAGTCCCGGACGCGCTCC 133  
DB 1991 TCCGCAACGCGCTCGACATCTGCTGCGCATCGCGCGGTACACGCGCTCCACAGGTGC 2050  
QY 134 CGATGATGCAGCGCGCGCGGTCAAGCTGTGTAAGACAGGCGCGCGCTACGCGCGCG 193  
DB 2051 CGATGATGCAGCGCGCGCATCGCGGTGTCGCGCATCGCGCGCGCGCTACGCGCGCG 2110  
QY 194 ACCACGCGCGCTCGACGATCTGCTGTGTAAGACCTCTACGACGTATCATACGAGATGT 253  
DB 2111 ACCACGCGCATCGACCGCGCTCTCTGTGCGACCTGTGACACGATCATCATCCGACT 2170  
QY 254 GCGCGCTGCAGGACCTGTGTAT 275  
DB 2171 GCGCGCTGCAGGACGAGTGTAT 2192

RESULT 7  
ABZ66707 standard; DNA; 1458 BP.  
XX  
AC ABZ66707;  
XX  
DT 21-MAR-2003 (first entry)  
XX  
DE Orthosomycin biosynthetic polynucleotide SEQ ID NO 76.  
XX  
KM Orthosomycin; biosynthesis; evernimycin; avilamycin; gene; ds.  
XX  
OS Streptomycetes mobaraensis.  
XX  
PN WO200279505-A2.  
XX  
PD 10-OCT-2002.  
XX  
PF 28-MAR-2002; 2002MO-CA000432.  
XX  
PR 28-MAR-2001; 2001US-0279095P.  
PR 30-MAR-2001; 2001US-0279709P.  
PR 20-APR-2001; 2001US-0285214P.  
XX  
PA (ECOP-) ECOPIA BIOSCIENCES INC.  
XX  
PI Farneet CM, Zazopoulos E, Staffa A;

XX  
DR WPI: 2003-058435/05.  
DR P-PSDB; ABP99244.  
XX  
PT Identifying orthosomycin biosynthetic gene, gene fragment or gene  
PT cluster, by detecting presence of nucleic acid sequence corresponding to  
PT 17 of flambamycins protein families.  
XX  
PS Example 2; Page 231-232; 51pp; English.  
XX  
CC The invention relates to identifying orthosomycin biosynthetic genes and  
CC its fragment/gene cluster (ABZ6670-ABZ66813), comprising detecting the  
CC presence of a nucleic acid sequence coding for a polypeptide (ABP99207-  
CC ABP99362). The method is useful for identifying an orthosomycin  
CC biosynthetic gene, gene fragment or gene cluster, especially an  
CC evernimycin-type or avilamycin-type orthosomycin biosynthetic gene,  
CC gene fragment or gene cluster. The method is useful for detecting the  
CC presence of any organism that contains DNA for the production of  
CC orthosomycins (both evernimycin-type orthosomycins and avilamycin-type  
CC orthosomycins) regardless of the level at which genes for orthosomycin  
CC production are expressed by the organism or the amount of orthosomycin  
CC produced by the organism. This allows for the detection of new  
CC orthosomycin natural products, not produced by the organism  
XX  
SQ Sequence 1458 BP; 229 A; 530 C; 491 G; 208 T; 0 U; 0 Other;  
Query Match 17.6%; Score 54.8; DB 10; Length 1458;  
Best Local Similarity 49.0%; Pred. No. 0.41;  
Matches 146; Conservative 0; Mismatches 152; Indels 0; Gaps 0;  
QY 6 CGAGCAGAAAGACCTGACGCGGCTGCGCGGAGCTCGAGCCCTCGACCGGAGCGCTCT 65  
DB 684 CACACGAGACCGCTTCTCTGAGCGCGCAGAGAGACAGACCTGAGGCGAGTCTCGC 743  
QY 66 GGAACAGGTGCGCGCGCGCATCGACTCGGTGTCGCGCATCGCGGTACAAGTCCCGGCA 125  
DB 744 CTGCGTGTCTCAAGAGAACCCGACCGTCCGCGCATCATGGCGAGGCGCATACCGCC 803  
QY 126 CGCGGTCCGATGATGACGCGCGCGGTGACCTGTGTCAGAGACAGGCGCGCGCTTA 185  
DB 804 GGGCGTCCGGGTGAGAGAGACTTCTCTACGTCTGCGCGACAGCTTCCACGCGCGCGCTA 863  
QY 186 CGCGCGCGACCAAGCGCTCGACGATCTGCTGTGTAAGACCTCTACGACGTATCATAC 245  
DB 864 CTACTCTGCGCGGAGCGCGCTGCTCTCTGACCGCGCTGTCTCAACCGCGCGCTCA 923  
QY 246 GGAATGTGCGCGGTGCAGGACCTGTGATGATGACCGGAGAGCGCTGACGCGCGAGAC 303  
DB 924 GCGCTCTACAGCGGAGCTGTGCGCGGAGCTGCTGCTGCTGCGGAGTGAACAGAGAC 981

RESULT 8  
ABZ66808/c  
ID ABZ66808 standard; DNA; 45055 BP.  
XX  
AC ABZ66808;  
XX  
DT 21-MAR-2003 (first entry)  
XX  
DE Orthosomycin biosynthetic gene cluster SEQ ID NO 277.  
XX  
KM Orthosomycin; biosynthesis; evernimycin; avilamycin; gene; ds.  
XX  
OS Streptomycetes mobaraensis.  
XX  
PN WO200279505-A2.  
XX  
PD 10-OCT-2002.  
XX  
PF 28-MAR-2002; 2002MO-CA000432.  
XX  
PR 28-MAR-2001; 2001US-0279095P.  
PR 30-MAR-2001; 2001US-0279709P.



```
PR 20-APR-2001; 2001US-0285214P.
XX
XX (ECOP-) ECOPIA BIOSCIENCES INC.
XX
XX Farnet CM, Zazopoulos E, Staffa A;
XX WPI; 2003-058435/05.
XX
XX Identifying orthosomycin biosynthetic gene, gene fragment or gene
XX cluster, by detecting presence of nucleic acid sequence corresponding to
XX 17 of flamamycins protein families.
XX
XX Example 2; Page 425-450; 511pp; English.
XX
XX The invention relates to identifying orthosomycin biosynthetic genes and
XX its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the
XX presence of a nucleic acid sequence coding for a polypeptide (ABP99207-
XX ABP99362). The method is useful for identifying an orthosomycin
XX biosynthetic gene, gene fragment or gene cluster, especially an
XX evernimicin-type or avilamycin-type orthosomycin biosynthetic gene,
XX gene fragment or gene cluster. The method is useful for detecting the
XX presence of any organism that contains DNA for the production of
XX orthosomycins (both evernimicin-type orthosomycins and avilamycin-type
XX orthosomycins) regardless of the level at which genes for orthosomycin
XX production are expressed by the organism or the amount of orthosomycin
XX produced by the organism. This allows for the detection of new
XX orthosomycin natural products, not produced by the organism
XX
XX Sequence 45055 BP; 6228 A; 16576 C; 16043 G; 6208 T; 0 U; 0 Other;
SQ
XX
XX Query Match 17.6%; Score 54.8; DB 10; Length 45055;
XX Best Local Similarity 49.0%; Pred. No. 0.3;
XX Matches 146; Conservative 0; Mismatches 152; Indels 0; Gaps 0;
XX
XX 6 CGAGCAGACGACCTGACGCGCTGCGCGGAGCTGACGCGCTGACGCGGACGCTCCT 65
XX 37629 CACCCAGACCGCTTCTGAGCGCCGCAAGAGACGACGACCTGAGCGCATGCTCGC 37570
XX
XX 66 GGACACGATGCGCGCGCGCATGCGCTGCTGCGCATGCGCGGATACAGTCCCGGCA 125
XX 37569 CTGCGTGTGTCAGAGAGAACCCGACGTCGCGCATGCTGCGGAGGCGAGTACCAAGC 37510
XX
XX 126 CGGCGTCCGATATGACGCGCGCGGCTGACGCTGCTGCAAGACAGGGCCCGGCTTA 185
XX 37509 GGGCGTCCGCGGTGAGCAGGACTTCTCTACGTCGCGCAGCTTCCACGCGCCCGGCTTA 37450
XX
XX 186 CGCGCGGACCAAGCGGCTGACGAATGCTTCCGTGTAACCTTACAGAGTGATCATCAC 245
XX 37449 CTACCTGCGCGGAGCGCGCTGCTTCTGACCGCGCTGCTTCAACCGCGCTTCACTT 37390
XX
XX 246 GGAGATGTGCGCGCTGAGAGACCTGATGATGAGCCGAGAGAGCTTACGCGCGAGGAC 303
XX 37389 GGGCGCTTACAGCGGAGATGTGCGCGGAGCGTCCGTGCTGCCACGCTGAACGAGAC 37332
XX
XX
XX RESULT 9
XX ABD12889
XX ID ABD12889 standard; DNA; 507 BP.
XX
XX ABD12889;
XX
XX 29-UTL-2004 (first entry)
XX
XX Pseudomonas aeruginosa polynucleotide #11493.
XX
XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
XX antibacterial.
XX
XX Pseudomonas aeruginosa.
XX
XX US6511795-B1.
XX
XX 22-APR-2003.
XX
XX
```

```
XX
XX 18-FEB-1999; 99US-00252991.
XX
XX 18-FEB-1998; 98US-0074788P.
XX
XX 27-UTL-1998; 98US-0094190P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX WPI; 2003-615309/58.
XX
XX P-PSDB; AB079318.
XX
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
XX useful as molecular targets for diagnostics, prophylaxis and treatment of
XX pathological conditions resulting from bacterial infection.
XX
XX Disclosure, SEQ ID NO 11493; 455pp; English.
XX
XX The invention relates to Pseudomonas aeruginosa polypeptides and the
XX polynucleotides encoding them. The sequences are useful in diagnosis and
XX therapy of pathological conditions, as molecular targets for diagnostics,
XX prophylaxis and treatment of pathological conditions resulting from a
XX bacterial infection, for evaluating a compound, such as a polypeptide,
XX for the ability to bind a P. aeruginosa nucleic acid, as components of
XX effective antibacterial targets, as targets for antibacterial drugs,
XX including anti-P. aeruginosa drugs, as templates for recombinant
XX production of P. aeruginosa-derived peptides or polypeptides, as target
XX components for diagnosis and/or treatment of P. aeruginosa-caused
XX infection, and in detection of P. aeruginosa sequences or other sequences
XX of Pseudomonas species using bioclip technology. Sequences ABD01197-
XX ABD1967 represent P. aeruginosa polynucleotides of the invention. Note:
XX The sequence data for this patent did not form part of the printed
XX specification but was obtained in electronic format from USPTO at
XX seqdata.uspto.gov/sequence.html
XX
XX Sequence 507 BP; 86 A; 176 C; 163 G; 82 T; 0 U; 0 Other;
SQ
XX
XX Query Match 17.5%; Score 54.6; DB 11; Length 507;
XX Best Local Similarity 49.8%; Pred. No. 0.49;
XX Matches 138; Conservative 0; Mismatches 139; Indels 0; Gaps 0;
XX
XX 33 CGGAGCTGACGCGCTTGAAGGAGCGCTCTGACACGCTGCGCGCATGCACT 92
XX 149 CGAAGGCTTCTTCTGAGCGCTGCGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 208
XX
XX 93 CGGTGCGCATGCGCGCTTGAAGGAGCGCTTGAAGGAGCGCTTGAAGGAGCGCTTGAAG 152
XX 209 CGCGTGAAGTGTGCTTGAATTTCAACGAGGCTGCGCGCGAGACCACTTGTCTTACCG 268
XX
XX 153 GGTGACCTGATGACAGGAGCGCGCTTGAAGGAGCGCTTGAAGGAGCGCTTGAAGGAG 212
XX 269 TCGGACCTGCTTCTGCGGCGCTTGAAGGAGCGCTTGAAGGAGCGCTTGAAGGAGGAT 328
XX
XX 213 GTTCTGTGAACCTTGAAGGAGCGCTTGAAGGAGCGCTTGAAGGAGCGCTTGAAGGAG 272
XX 329 GGGCAGAGAGGCGCTGACCTGCAATTCAACAGCGACATGCGCCGATGACAAACAGCGC 388
XX
XX 273 GATGAGCGCGGAGAGCGCTTGAAGGAGCGCGCGCGG 309
XX 389 CGATGGACGCTGCGCGGAGCGCTTGAAGGAGCGCGG 425
XX
XX
XX RESULT 10
XX ABD12819/C
XX ID ABD12819 standard; DNA; 1359 BP.
XX
XX ABD12819;
XX
XX 29-UTL-2004 (first entry)
XX
XX Pseudomonas aeruginosa polynucleotide #11423.
XX
XX
```



KM Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
KM antibacterial.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN US6551795-B1.  
XX  
PD 22-APR-2003.  
XX  
PF 18-FEB-1999; 99US-00252991.  
XX  
PR 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX  
DR WPI; 2003-615309/58.  
DR P-PSDB; ABO79248.  
XX  
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
PS Disclosure; SEQ ID NO 11423; 455pp; English.  
XX  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
XX polynucleotides encoding them. The sequences are useful in diagnosis and  
XX therapy of pathological conditions, as molecular targets for diagnostics,  
XX prophylaxis and treatment of pathological conditions resulting from a  
XX bacterial infection, for evaluating a compound, such as a polypeptide,  
XX for the ability to bind a P. aeruginosa nucleic acid, as components of  
XX effective antibacterial targets, as targets for antibacterial drugs,  
XX including anti-P. aeruginosa drugs, as templates for recombinant  
XX production of P. aeruginosa-derived peptides or polypeptides, as target  
XX components for diagnosis and/or treatment of P. aeruginosa-caused  
XX infection, and in detection of P. aeruginosa sequences or other sequences  
XX of Pseudomonas species using biochip technology. Sequences ABD01397-  
XX ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
XX The sequence data for this patent did not form part of the printed  
XX specification but was obtained in electronic format from USPTO at  
XX seqdata.uspto.gov/sequence.html  
XX  
SQ Sequence 1359 BP; 212 A; 463 C; 446 G; 238 T; 0 U; 0 Other;  
XX  
Query Match 17.5%; Score 54.6; DB 11; Length 1359;  
Best Local Similarity 49.8%; Pred. No. 0.45;  
Matches 138; Conservative 0; Mismatches 139; Indels 0; Gaps 0;  
XX  
QY 33 CGCGAGCTCGACGCCCTCGACGGGACGCTCTGACACGGTGGCGCGCATCGACT 92  
Db 928 CGAAGCGTTCTTCTCTCGAGCGCTGCGCGCGCATGCTGCTGCTGCGCGCGCTACAT 869  
XX  
QY 93 CGGTGTCGCGATCGCGCGGTACAGTCCCGGACGGGCTGCCATGATGACGCCGCCG 152  
Db 868 CGCGGTGAGTTCCTCGATCTTCAACGGGCTTGGCGCGACCACTTGTCTTACCG 809  
XX  
QY 153 GGTACGCTGTGTAAGACAGGCGCGCGCTACGCGCGACCAAGCGCTCGACGATC 212  
Db 808 TCGGACCTGTCTCTCGCGCGCTTGCACCGAGCTCCGCGAGATCTGCGCGCAATT 749  
XX  
QY 213 GTTCTGTGTAACCTTACGACGTGATCATCAGGAGATGTCCGCTCGAGACCTGT 272  
Db 748 GGGCAAGAGGCGCTGACCTGACCTGACATTCAACAGGACATCGCCGTATCGCAAAAGCG 689  
XX  
QY 273 GATGAGCGGGAGAGCTGACGCGCGCGGACCGCGCG 309  
Db 688 CGATGGACCTGGCGCGACCTTCAAGAGCGCGCG 652  
XX  
RESULT 11  
ABD12866

ID ABD12866 standard; DNA; 1371 BP.  
XX  
AC ABD12866;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Pseudomonas aeruginosa polynucleotide #11470.  
XX  
KM Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
KM antibacterial.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN US6551795-B1.  
XX  
PD 22-APR-2003.  
XX  
PF 18-FEB-1999; 99US-00252991.  
XX  
PR 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX  
DR WPI; 2003-615309/58.  
DR P-PSDB; ABO79295.  
XX  
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
PS Disclosure; SEQ ID NO 11470; 455pp; English.  
XX  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
XX polynucleotides encoding them. The sequences are useful in diagnosis and  
XX therapy of pathological conditions, as molecular targets for diagnostics,  
XX prophylaxis and treatment of pathological conditions resulting from a  
XX bacterial infection, for evaluating a compound, such as a polypeptide,  
XX for the ability to bind a P. aeruginosa nucleic acid, as components of  
XX effective antibacterial targets, as targets for antibacterial drugs,  
XX including anti-P. aeruginosa drugs, as templates for recombinant  
XX production of P. aeruginosa-derived peptides or polypeptides, as target  
XX components for diagnosis and/or treatment of P. aeruginosa-caused  
XX infection, and in detection of P. aeruginosa sequences or other sequences  
XX of Pseudomonas species using biochip technology. Sequences ABD01397-  
XX ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
XX The sequence data for this patent did not form part of the printed  
XX specification but was obtained in electronic format from USPTO at  
XX seqdata.uspto.gov/sequence.html  
XX  
SQ Sequence 1371 BP; 241 A; 445 C; 464 G; 421 T; 0 U; 0 Other;  
XX  
Query Match 17.5%; Score 54.6; DB 11; Length 1371;  
Best Local Similarity 49.8%; Pred. No. 0.45;  
Matches 138; Conservative 0; Mismatches 139; Indels 0; Gaps 0;  
XX  
QY 33 CGCGAGCTCGACGCCCTCGACGGGACGCTCTGACACGGTGGCGCGCATCGACT 92  
Db 486 CGAAGCGTTCTTCTCTCGAGCGCTGCGCGCGCATGCTGCTGCTGCGCGCGCTACAT 545  
XX  
QY 93 CGGTGTCGCGATCGCGCGGTACAGTCCCGGACGGGCTCCGATGATGACGCCGCCG 152  
Db 546 CGCGGTGAGTTCCTCGATCTTCAACGGGCTTGGCGCGACCACTTGTCTTACCG 605  
XX  
QY 153 GGTACGCTGTGTAAGACAGGCGCGCGCTACGCGCGACCAAGCGCTCGACGATC 212  
Db 606 TCGGACCTGTCTCTCGCGCGCTTGCACCGAGCTCCGCGAGATCTGCGCGCAATT 665  
XX  
QY 213 GTTCTGTGTAACCTTACGACGTGATCATCAGGAGATGTCCGCTCGAGACCTGT 272  
Db 666 GGGCAAGAGGCGCTGACCTGACGATCTTCAACAGGACGATCGCCGTATCGCAAAAGCG 725



Qy	273	GATGAGCCGGAGAGCCTTACGCGCCAGACCGCGG	309
Db	726	CGATGGCAGCCTTGCGCGCGACCTCTCAAGACGCGCGG	762
RESULT 12			
ID	AD573531	standard; cDNA; 73882 BP.	
XX	AD573531;		
XX	02-DEC-2004	(first entry)	
DE	tcp gene cluster.		
XX	ss; gene; tcp ; gene cluster ; biosynthesis ; glycopeptide ; teicoplanin;		
KM	antibiotic; bal; cep; com ; sta ; 4-hydroxy-phenylglycine ; 3;		
KM	5-dihydroxy-phenylglycine; heptapeptide skeleton ; cross-linking ;		
KM	aromatic residue; N-acyl beta-D-glucosamine ; core structure ; export;		
KM	resistance; chloramine; alpha-hydroxylation ; mannosyl attachment.		
XX	Actinoplanes teichomyceticus; ATCC31121.		
OS			
FH	Key	Location/Qualifiers	
FT	CDS	652..1995	
FT		/*tag= a	
FT		/label= ORF1	
FT		/note= "Required for export of teicoplanin outside of the	
FT		cytoplasm, and for conferring resistance to teicoplanin"	
FT		/product= "MurF"	
FT		2017..3234	
FT	CDS	/*tag= b	
FT		/label= ORF2	
FT		/note= "Required for export of teicoplanin outside of the	
FT		cytoplasm, and for conferring resistance to teicoplanin"	
FT		3227..4264	
FT	CDS	/*tag= c	
FT		/label= ORF3	
FT		/note= "Required for export of teicoplanin outside of the	
FT		cytoplasm, and for conferring resistance to teicoplanin"	
FT		4264..4872	
FT		/*tag= d	
FT		/label= ORF4	
FT		/note= "Required for export of teicoplanin outside of the	
FT		cytoplasm, and for conferring resistance to teicoplanin"	
FT		complement(5634..6173)	
FT	CDS	/*tag= e	
FT		/label= ORF5	
FT		/note= "Required for export of teicoplanin outside of the	
FT		cytoplasm, and for conferring resistance to teicoplanin"	
FT		complement(6286..7377)	
FT	CDS	/*tag= f	
FT		/label= ORF6	
FT		/note= "Required for export of teicoplanin outside of the	
FT		cytoplasm, and for conferring resistance to teicoplanin	
FT		and required for regulating expression of one or more of	
FT		the genes of tcp gene cluster"	
FT		complement(7385..8065)	
FT	CDS	/*tag= g	
FT		/label= ORF7	
FT		/note= "Required for regulating expression of one or more	
FT		of the genes of tcp gene cluster"	
FT		8924..10105	
FT	CDS	/*tag= h	
FT		/label= ORF8	
FT		/note= "Required for the addition and formation of N-acyl	
FT		-beta-glucosamine residues"	
FT		10597..16827	
FT		/*tag= i	
FT		/label= ORF9	
FT		/note= "Required for synthesis of the heptapeptide	
FT		skeleton"	
FT	CDS		

FT	16824. .20009	
FT	/tag= j	
FT	/label= ORF10	
FT	/note= "Required for synthesis of the heptapeptide skeleton"	
CDS	20053. .32256	
FT	/tag= k	
FT	/label= ORF11	
FT	/note= "Required for synthesis of the heptapeptide skeleton"	
CDS	32276. .37873	
FT	/tag= l	
FT	/label= ORF12	
FT	/note= "Required for synthesis of the heptapeptide skeleton"	
CDS	37886. .38095	
FT	/tag= m	
FT	/label= ORF13	
FT	38401. .39222	
FT	/tag= n	
CDS	/label= ORF14	
FT	39268. .41049	
FT	/tag= o	
FT	/label= ORF15	
CDS	/note= "Required for attachment of mannosyl residue"	
FT	41343. .43289	
FT	/tag= p	
FT	/label= ORF16	
FT	/note= "Required for export of teicoplanin outside of the cytoplasm, and for conferring resistance to teicoplanin"	
CDS	43373. .43585	
FT	/tag= q	
FT	/label= ORF17	
CDS	43696. .44871	
FT	/tag= r	
FT	/label= ORF18	
FT	/note= "Required for cross-linking of aromatic residues atposition 2 and 4, 4 and 6, 1 and 3, and 5 and 7 of teicoplanin"	
CDS	44894. .46048	
FT	/tag= s	
FT	/label= ORF19	
FT	/note= "Required for cross-linking of aromatic residues atposition 2 and 4, 4 and 6, 1 and 3, and 5 and 7 of teicoplanin"	
CDS	46038. .47234	
FT	/tag= t	
FT	/label= ORF20	
FT	/note= "Required for cross-linking of aromatic residues atposition 2 and 4, 4 and 6, 1 and 3, and 5 and 7 of teicoplanin"	
CDS	47484. .49004	
FT	/tag= u	
FT	/label= ORF21	
FT	/note= "Required for the chlorination of the aromatic residues"	
CDS	49166. .50344	
FT	/tag= v	
FT	/label= ORF22	
FT	/note= "Required for cross-linking of aromatic residues atposition 2 and 4, 4 and 6, 1 and 3, and 5 and 7 of teicoplanin"	
CDS	50461. .51687	
FT	/tag= w	
FT	/label= ORF23	
FT	/note= "Required for the addition and formation of N-acyl-beta-glucosamine residues"	
CDS	51833. .52804	
FT	/tag= x	
FT	/label= ORF24	
FT	/note= "Required for the addition and formation of N-acyl-beta-glucosamine residues"	
CDS	52948. .54540	
FT		



FT	/+cag= y
FT	/label= ORF25
FT	/note= "Required for the beta-hydroxylation of the Tyr residues"
FT	54757. .56553
FT	/+cag= z
FT	/label= ORF26
FT	/note= "Required for the addition and formation of N-acyl-beta-glucosamine residues"
FT	56924. .57979
FT	/+cag= aa
FT	/label= ORF27
FT	58746. .59735
FT	/+cag= ab
FT	/label= ORF28
FT	/note= "Required for regulatingexpression of one or more of the genes of tcp gene cluster"
FT	60477. .62855
FT	/+cag= ac
FT	/label= ORF29
FT	/note= "Required for regulatingexpression of one or more of the genes of tcp gene cluster"
FT	63036. .64157
FT	/+cag= ad
FT	/label= ORF30
FT	/note= "Required for synthesis of the 3,5-dihydroxy-phenylglycine residues"
FT	64154. .64819
FT	/+cag= ae
FT	/label= ORF31
FT	/note= "Required for synthesis of the 3,5-dihydroxy-phenylglycine residues"
FT	64816. .66117
FT	/+cag= af
FT	/label= ORF32
FT	/note= "Required for synthesis of the 3,5-dihydroxy-phenylglycine residues"
FT	66146. .66952
FT	/+cag= ag
FT	/label= ORF33
FT	/note= "Required for synthesis of the 3,5-dihydroxy-phenylglycine residues"
FT	66991. .68352
FT	/+cag= ah
FT	/label= ORF34
FT	/note= "Required for export of teicoplanin outside of the cytoplasm, and for conferring resistance to teicoplanin"
FT	68349. .68945
FT	/+cag= ai
FT	/label= ORF35
FT	69083. .70180
FT	/+cag= aj
FT	/label= ORF36
FT	/note= "Required for synthesis of the 4-hydroxy-phenylglycine residues, and required for synthesis of the 3,5-dihydroxy-phenylglycine residues"
FT	70266. .71321
FT	/+cag= ak
FT	/label= ORF37
FT	/note= "Required for synthesis of the 4-hydroxy-phenylglycine residues"
FT	71318. .72412
CDS	

Db	29773	CGCTGTGTGGGTTACCCGCCGCCAGCGCTCAGATGAACAACCTCTTCACAGGGGATC	29833
QY	121	CGGCACGGCGCTCCCGATGATGCAAGCCGGGCGGGTTCAGCTGTGTCAAGCAGGCGCCG	180
Db	29833	TGTGCTGTGTCTGTGCGCGGCTGACCGGCGCCGACGAGTGTGTCTTCGGTTCGCCGCTG	29892
QY	181	CGGTACGGCGCGACACGCGCTGCAGCAATGGTTCCTGTGTAACTCTACGACGATATC	240
Db	29893	GCCGGGCGGCCCGCCGAGATCCCGCGCTCGGATCCGTGTCTGGGCTGTTCATGAACACG	29952
QY	241	ATCAGCAGAGATGTGCCGCTGCAGACCTGTGTATGAGCCGGGAGAGGCTGACGGCCGAG	300
Db	29953	CTGCCGATCCGGGTGTGGCTTCAGACGGCGCCGAACGTTCTCTGACATGTCTACACGACTG	3012
QY	301	GACCGGCGG	309
Db	30013	CAGCGGCGG	30021

ID	AA064204	standard	CDNA	1208 BP.
XX	AA064204			
AC				
XX				
DT	25-MAR-2003	(revised)		
DT	18-NOV-1994	(first entry)		
XX				
DE	snac gene encoding enzyme in streptogramin biosynthetic pathway.			
XX				
XX	Antibiotic; streptogramin; snac; snab; snac; biosynthesis; enzyme;			
KM	biosynthetic pathway; Streptomyces pristinaespiralis; da.			
KW				
XX	Streptomyces pristinaespiralis.			
OS				
XX				
XX	Key	Location/Qualifiers		
XX	Key	1..1209		
XX	CDS	/*tag= a		
XX				
XX	FR6696189-A1.			
XX				
XX	01-APR-1994.			
XX				
XX	25-SEP-1992; 92FR-00011441.			
XX				
XX	25-SEP-1992; 92FR-00011441.			
XX				
XX	(RHON ) RHONE POULENC RORER SA.			
XX				
XX	Blanc V, Blanche F, Cruzet J, Jacques N, Lacroix P, Thibaut D;			
XX	Zagorec M;			
XX				
XX	WPI; 1994-128286/16.			
XX				
XX	P-PSDB; AARS4204.			
XX				
XX	DNA involved in streptogramin antibiotic biosynthesis - for prodn. or bida			
XX	-conversion of streptogramin(s) or prodn. of streptogramin intermediates.			
XX	derivs. or hybrid antibiotics.			
XX				
XX	Claim 2; Page 54-55; 83pp; French.			
XX				
XX	The snac gene product is involved in the biosynthesis of streptogramins,			
XX	antibiotics active against Gram-positive bacteria. The identification of			
XX	the sequences encoding the enzymes involved in the biosynthetic pathway			
XX	means that they can be isolated and manipulated. Mutant microorganisms in			
XX	which a step in the streptogramin biosynthetic pathway is blocked can be			
XX	cultured to produce streptogramin intermediates, which may later be			
XX	converted to streptogramin derivatives. Recombinant cells may also be			
XX	used for the bioconversion of streptogramins from one form to another or			
XX	for the production of hybrid antibiotics. (Updated on 25-MAR-2003 to			
XX	correct PN field.)			
XX				
XX	Sequence 1208 BP; 190 A; 500 C; 360 G; 158 T; 0 U; 0 Other;			



Query Match 17.3%; Score 54; DB 2; Length 1208;  
Best Local Similarity 50.0%; Pred. No. 0.58;  
Matches 135; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 38 AGCTCAGCGCCCTTCGACGGAGCGCTCTGACACGGTGGCGGCCGATGACCTGGTG 97  
DB 860 AGGTTCGACCGCTTCGCGCGCTACGCGATGCGTGGTGGCAAGAAAGTGTGCGCGCG 919  
QY 98 TCCGATCGCGCGGTGTAACAATCCCGGACGGCGCTCCGATGATGACGCCGCGCGGTCA 157  
DB 920 GCTTCGCTTCCTCCGCTGCGAGGTCAAGTCCCTACGCTACGCGAAGCGACCGGTG 979  
QY 158 GCCTGTGCAAGAGCAGGGCGCGCGCTACGCGCGCGACGCGCTTCGACGAATGCTTC 217  
DB 980 GCTGTGTCGTGAGAGCTTGGGACCGGCAACCGTCCGCCAGAGCGCATGAGAAAGGCCA 1039  
QY 218 TGTGTAACCTCTACGACGTATCATCAGGAGATGTGCCGCTTCGAGAGACTGTGTGA 277  
DB 1040 TCACCGAGGTCTTCGACCTGCGCGCGCGCGCATTCATCCGCGACTTCGCTGTGGGC 1099  
QY 278 GCGGGAGAGCTGACGCGCGCGAGAACCGGC 307  
DB 1100 CCATCTACGCGCGCACCGCGCTACGGCC 1129

RESULT 14  
AA064201  
ID AA064201 standard; cDNA; 5392 BP.  
AC AA064201;  
XX  
DT 25-MAR-2003 (revised)  
DT 18-NOV-1994 (first entry)  
XX  
DE Sequence comprising the *snaA*, *snaB* and *snaC* gene cluster.  
XX  
KM Antibiotic; streptogramin; *snaA*; *snaB*; *snaC*; biosynthesis; enzyme;  
KM biosynthetic pathway; Streptomyces pristinaespiralis; ds.  
XX  
OS Streptomyces pristinaespiralis.  
XX  
XX  
PN FR2696189-A1.  
XX  
PD 01-APR-1994.  
XX  
PF 25-SEP-1992; 92FR-00011441.  
XX  
PR 25-SEP-1992; 92FR-00011441.  
XX  
PA (RHON) RHONE-POULENC RORER SA.  
XX  
PI Blanc V, Blanche F, Crouzet J, Jacques N, Lacroix P, Thibaut D;  
PI Zagorec M;  
XX  
DR WPI; 1994-128286/16.  
XX  
PT DNA involved in streptogramin antibiotic biosynthesis - for prodn. or bio-  
PT -conversion of streptogramin(s) or prodn. of streptogramin intermediates,  
PT derivs. or hybrid antibiotics.  
XX  
PS Disclosure; Page 44-47; 83pp; French.  
XX  
CC This sequence comprises the *snaA*, *snaB* and *snaC* genes which are involved  
CC in the biosynthesis of streptogramins, antibiotics active against Gram-  
CC positive bacteria. The identification of the sequences encoding the  
CC enzymes involved in the biosynthetic pathway means that they can be  
CC isolated and manipulated. Mutant microorganisms in which a step in the  
CC streptogramin biosynthetic pathway is blocked can be cultured to produce  
CC streptogramin intermediates, which may later be converted to  
CC streptogramin derivatives. Recombinant cells may also be used for the  
CC bioconversion of streptogramins from one form to another or for the  
CC production of hybrid antibiotics. (Updated on 25-MAR-2003 to correct PN

CC field.)  
XX  
SQ Sequence 5392 BP; 811 A; 2161 C; 1671 G; 749 T; 0 U; 0 Other;  
QY  
Best Local Similarity 17.3%; Score 54; DB 2; Length 5392;  
Best Local Similarity 50.0%; Pred. No. 0.51;  
Matches 135; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 38 AGCTCAGCGCCCTTCGACGGAGCGCTCTGACACGGTGGCGGCCGATGACCTGGTG 97  
DB 4417 AGGTTCGACCGCTTCGCGCGCTACGCGATGCGTGGTGGCAAGAAAGTGTGCGCGCG 4476  
QY 98 TCCGATCGCGCGGTGTAACAATCCCGGACGGCGCTCCGATGATGACGCCGCGCGGTCA 157  
DB 4477 GCTTCGCTTCCTCCGCTGCGAGGTCAAGTCCCTACGCTACGCGAAGCGACCGGTG 4536  
QY 158 GCCTGTGCAAGAGCAGGGCGCGCGCTACGCGCGCGACGCGCTTCGACGAATGCTTC 217  
DB 4537 GCTGTGTCGTGAGAGCTTGGGACCGGCAACCGTCCGCCAGAGCGCATGAGAAAGGCCA 4596  
QY 218 TGTGTAACCTCTACGACGTATCATCAGGAGATGTGCCGCTTCGAGAGACTGTGTGA 277  
DB 4597 TCACCGAGGTCTTCGACCTGCGCGCGCGCGCATTCATCCGCGACTTCGCTGGGC 4656  
QY 278 GCGGGAGAGCTGACGCGCGCGAGAACCGGC 307  
DB 4657 CCATCTACGCGCGCACCGCGCTACGGCC 4686

RESULT 15  
AB266725  
ID AB266725 standard; DNA; 837 BP.  
AC AB266725;  
XX  
DT 21-MAR-2003 (first entry)  
XX  
DE Orthosomycin biosynthetic polynucleotide SEQ ID NO 112.  
XX  
KM Orthosomycin; biosynthesis; everninomycin; avilamycin; gene; ds.  
XX  
OS Streptomyces mobaraensis.  
XX  
PN WO200279505-A2.  
XX  
PD 10-OCT-2002.  
XX  
PF 28-MAR-2002; 2002WO-CA000432.  
XX  
PR 28-MAR-2001; 2001US-0279095P.  
PR 30-MAR-2001; 2001US-0279709P.  
PR 20-APR-2001; 2001US-0285214P.  
XX  
PA (ECOP) ECOPIA BIOSCIENCES INC.  
XX  
PI Farnet CM, Zazopoulos E, Staffa A;  
PI  
XX  
DR WPI; 2003-058435/05.  
DR P-PDB; ABP99262.  
XX  
PT Identifying orthosomycin biosynthetic gene, gene fragment or gene  
PT cluster, by detecting presence of nucleic acid sequence corresponding to  
PT 17 of flambamycins protein families.  
XX  
PS Claim 12; Page 266; 51pp; English.  
XX  
CC The invention relates to identifying orthosomycin biosynthetic genes and  
CC its fragment/gene cluster (AB266670-AB2666813), comprising detecting the  
CC presence of a nucleic acid sequence coding for a polypeptide (ABP99207-  
CC ABP99362). The method is useful for identifying an orthosomycin  
CC biosynthetic gene, gene fragment or gene cluster, especially an  
CC everninomycin-type or avilamycin-type orthosomycin biosynthetic gene,  
CC gene fragment or gene cluster. The method is useful for detecting the



CC presence of any organism that contains DNA for the production of  
 CC orthosomycins (both everimomycin-type orthosomycins and avilamycin-type  
 CC orthosomycins) regardless of the level at which genes for orthosomycin  
 CC production are expressed by the organism or the amount of orthosomycin  
 CC produced by the organism. This allows for the detection of new  
 CC orthosomycin natural products, not produced by the organism  
 XX

SO Sequence 837 BP; 105 A; 329 C; 308 G; 95 T; 0 U; 0 Other;

Query Match 17.2%; Score 53.6; DB 10; Length 837;

Best Local Similarity 52.1%; Pred. No. 0.71;

Matches 147; Conservative 0; Mismatches 129; Indels 6; Gaps 1;

Qy	28	CTGGCGCGGAGCTCGACGCGCTTCGAGCGGACGCTCTGACACAGGTGCGCGCGCATC	87
Db	151	CTGCTCGACCTCCGCGACGCGCTGCGCGAGAGCTCCGCGGCGCTCCGCGCGCTGCTC	210
Qy	88	GACCTCGGTGTCGCGATCGCGGCTACAGTCCCGACGCGCTCCGATGATGACGCC	147
Db	211	GACTTCGCGCGCGCGACGTGCGCTTACCGCGCGCTGCTCCGCGCGCGCGACGTGAGACC	270
Qy	148	GCGCGGCTCAGCTGTCAAGACAGGCGCGCGCTACGCGCGCGACACAGGCGCTGAC	207
Db	271	GCGAGATGCGCGCGCGCGAGACCTGACCGCGACGAGCTGACGCGGAGCGGCGCTG	330
Qy	208	GAATCGTTCGTGTGAACCTCTACGAGCTGATCATACGAGATGTGCGGTGAGAC	267
Db	331	AGCGCCCTGCGCGCGCGCTGCTGACGCGATCTCTC-----CACCAGGTCTTGGAG	384
Qy	268	CTGGTATGAGCGCGAGAGCTTGACGCGCGAGACCGCGCG	309
Db	385	CACGTGACGAGCCGAGACCCACCTCGCGGAGGCGCTCCGG	426

Search completed: October 4, 2005, 21:38:06  
 Job time : 191.488 secs



This Page Blank (uspio)







```
Db      299 GCCGCTCGAGAGCAGGTGAT 320

RESULT 2
US-09-987-614A-4
; Sequence 4, Application US/09987614A
; Patent No. 683382
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSETT, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/09/987,614A
; PRIORITY FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US/08/765,907
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 387
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-4

Query Match      43.5%; Score 135.6; DB 4; Length 387;
Best Local Similarity 69.8%; Pred. No. 6.3e-21;
Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY      14 ACGAGCTGCAGCGCGCTGCGCGGAGCTCGACGCCCTCGACGGAGACGCTCTTGAGACAGG 73
      59 ACCCCCTCGACGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTTGAGACAGG 118
      74 TCGGCGCGCGCGCATCGACCTTGTGTCGCGCATCGCGCGGTACAAAGTCCCGGACGCGGTCC 133
      119 TCGGCACACCGCTCGACATCTGCTGCGCATCGCGCGGTACAAAGTCCCGGTCCACAGGTGC 178
      134 CGATGATGCAGCGCGCGCGGTGCACTGCTCAAGAGCAGAGCGCGCGCGGTACGCGCGCG 193
      179 CGATGATGCAGCGCGCGCGGTGCACTGCTCAAGAGCAGAGCGCGCGCGGTACGCGCGCG 238
      194 ACCAGCGCGCTCGACGATCGTCTGTGATGATCTTACGAGTATCATCAGGAGATGT 253
      239 ACCAGCGCATCGACCGCGCTTCTCTGCGCATCTGTACAGACATCATACGAGACT 298
      254 GCCGCTCGAGAGCAGTGTGAT 275
      299 GCCGCTCGAGAGCAGTGTGAT 320

RESULT 3
US-08-765-907A-1
; Sequence 1, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSETT, Gilles
```

```
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/08/765,907A
; PRIORITY FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2888
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-1

Query Match      43.5%; Score 135.6; DB 3; Length 2888;
Best Local Similarity 69.8%; Pred. No. 7.5e-21;
Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY      14 ACGAGCTGCAGCGCGCTGCGCGGAGCTCGACGCCCTCGACGGAGACGCTCTTGAGACAGG 73
      1931 ACCCCCTCGACGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTTGAGACAGG 1990
      74 TCGGCGCGCGCGCATCGACCTTGTGTCGCGCATCGCGCGGTACAAAGTCCCGGACGCGGTCC 133
      1991 TCGGCACACCGCTCGACATCTGCTGCGCATCGCGCGGTACAAAGTCCCGGTCCACAGGTGC 2050
      134 TCGGCGCGCGCGCGCGCGCGGTGCACTGCTCAAGAGCAGAGCGCGCGCGGTACGCGCGCG 193
      2051 CGATGATGCAGCGCGCGCGGTGCACTGCTCAAGAGCAGAGCGCGCGCGGTACGCGCGCG 2110
      194 ACCAGCGCGCTCGACGATCGTCTGTGATGATCTTACGAGTATCATCAGGAGATGT 253
      2111 ACCAGCGCATCGACCGCGCTTCTCTGCGCATCTGTACAGACATCATCAGGAGACT 2170
      254 GCCGCTCGAGAGCAGTGTGAT 275
      2171 GCCGCTCGAGAGCAGTGTGAT 2192

RESULT 4
US-09-987-614A-1
; Sequence 1, Application US/09987614A
; Patent No. 683382
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSETT, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/09/987,614A
; PRIORITY FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US/08/765,907
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2888
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-1

Query Match      43.5%; Score 135.6; DB 4; Length 2888;
Best Local Similarity 69.8%; Pred. No. 7.5e-21;
Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY      14 ACGAGCTGCAGCGCGCTGCGCGGAGCTCGACGCCCTCGACGGAGACGCTCTTGAGACAGG 73
```



```
Db      1931 ACCCCCTGCAGCGGCTGCGCGCCCGCTGCAGCCCGCGAGCGCCCGCTCTGTGAGCCG 1990
      74 TCGCGCGCGCGCATGCACCTTCGTGTCCGATCGCGCGGTACAAGTCCCGGACGCGCTCC 133
Db      1991 TCGGCACACGCGCTGCACATCTGCTGCGATCGCGGAGTACAAGCGCGCTCCACAGGTGC 2050
      134 CGATGATGCAGCCCGCGCGCGGTGAGCTGTGATCAAGACAGAGCGCGCGCTTACGCGCGCG 193
Db      2051 CGATGATGCAGCGCCCGCGCGCGATGCGCCAGGTCCAGCCAGCGCGCGCGCTTACGCGCGCG 2110
      194 ACCAGCGCGCTGCAGCATGCTTCTCTGTGAACCTCTACGAGTATCATCAGGAGATGT 253
Db      2111 ACCAGCGCGCATGCAGCCCGCGCTTCTCTGCGACCTGTATCAGACAGATCATCAGGAACT 2170
      254 GCGCGCTGCAGAGCATGCTGTAT 275
      2171 GCGCGCTGCAGAGCATGCTGTAT 2192
```

RESULT 5  
US-09-252-991A-11493  
Sequence 11493, Application US/09252991A  
Patent No. 6551795

GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 11493  
LENGTH: 507  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11493

Query Match 17.5%; Score 54.6; DB 4; Length 507;  
Best Local Similarity 49.8%; Pred. No. 0.0029; Indels 0; Gaps 0;  
Matches 138; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```
QY      33 CGCGAGCTGCAGCGCCCTGCAGCGGACGCTCTGACACGCGTCCGCGCGCATGCACCT 92
      149 CGAAGCGTTCTTCTCTGAGCGCTGCGCGCGCGCATGCTGTGTGCGCGCGGCTACAT 208
      93 CGGTGTCGCGATCGCGCGGTACAAGTCCCGGACGCGCGTCCGATGATGACGCCGCGCG 152
Db      209 CGCGGTGAGTTGCGCTCGATCTTCAACGCGGCTCGGCGCGGACGACCACTTGCTTACCG 268
      153 GGTACACCTGTGTAAAGACAGGCGCGCGCTACGCGCGCGACGACGCGCTCGACGATC 212
Db      269 TCGCGACCTGTTCTCTGCGCGCTTTCGACCGGACGCTCCGAGCATCTGCGCAGCAATT 328
      213 GTTCTGTGTGAACCTCTACAGCATGATCATCAGGAGTGTGCGCGTGCAGGACCTGCT 272
Db      329 GGGCAAGAGGCGCTGAGCTTGCAGTTTCAACAGGACGATGCGCCGCTATGCAAAAGCG 388
      273 GATGACCGCGGAGAGCTTGAAGCGCGGACGAGACCGGCGCG 309
      389 CGATGCGAGCGCTGCGCGCGACCTCAAGAGCGGCGCG 425
```

RESULT 6  
US-09-252-991A-11423/c  
Sequence 11423, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 11423  
LENGTH: 1359  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11423

Query Match 17.5%; Score 54.6; DB 4; Length 1359;  
Best Local Similarity 49.8%; Pred. No. 0.0031; Indels 0; Gaps 0;  
Matches 138; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```
QY      33 CGCGAGCTGCAGCGCCCTGCAGCGGACGCTCTGACACGCGTCCGCGCGCATGCACCT 92
      928 CGAAGCGTTCTTCTCTGAGCGCTGCGCGCGCATGCTGTGTGCGCGCGGCTTACAT 869
      93 CGGTGTCGCGATCGCGCGGTACAAGTCCCGGACGCGCGTCCGATGATGACGCCGCGCG 152
Db      868 CGCGGTGAGTTGCGCTCGATCTTCAAGCGGCTCGGCGCGGACACACCTTGTCTTACCG 809
      153 GGTACACCTGTGTAAAGACAGGCGCGCGCTTACGCGCGCGACGACGCGCTCGACGATC 212
Db      808 TCGCGACCTGTTCTCTGCGCGCTTTCGACCGGACGCGTCCGAGCATCTGCGCAGCAATT 749
      213 GTTCTGTGTGAACCTCTACAGCATGATCATCAGGAGTGTGCGCGTGCAGGACCTGCT 272
Db      748 GGGCAAGAGGCGCTGAGCTTGCAGTTTCAACGCGACGATGCGCCGCTATGCAAAAGCG 689
      273 GATGACCGCGGAGAGCTTGAAGCGCGGACGAGACCGGCGCG 309
      688 CGATGCGAGCTGCGCGCGACCTCAAGAGCGGCGCG 652
```

RESULT 7  
US-09-252-991A-11470  
Sequence 11470, Application US/09252991A  
Patent No. 6551795

GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 11470  
LENGTH: 1371  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-11470

Query Match 17.5%; Score 54.6; DB 4; Length 1371;  
Best Local Similarity 49.8%; Pred. No. 0.0031; Indels 0; Gaps 0;  
Matches 138; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```
QY      33 CGCGAGCTGCAGCGCCCTGCAGCGGACGCTCTGACACGCGTCCGCGCGCATGCACCT 92
      486 CGAAGCGTTCTTCTCTGAGCGCTGCGCGCGCATGCTGTGTGCGCGCGGCTTACAT 545
      93 CGGTGTCGCGATCGCGCGGTACAAGTCCCGGACGCGCGTCCGATGATGACGCCGCGCG 152
```



Db 546 CGCGCTGAGTTTCCTCATCTTCAACGGGCTCGCGCCGAGACCACTTGTCTTACCG 605  
Qy 153 GGTACAGCTGTGTACAGAGAGGCGCCGCTTACGCGCGGACCAAGGCTTCAGCAATC 212  
Db 606 TGGGACCTGTGTCTTGGGGGCTTTCGACCGGACGCTTCGCGACATCTCGCGACGATTT 665  
Qy 213 GTTCTGTGTAACTTACGACGTATCATACGAGATGTGCGCGTTCGAGGACCTGT 272  
Db 666 GGGGCAAGAGGGCTTCGACCTGAGTTCAACAGCGACATCGCCGTATCGACAAAGAGG 725  
Qy 273 GATGAGCGGAGAGGCTTCGACGCGCGGAGGACCGGCGG 309  
Db 726 CGATGCGAGCTTGCGCCGACCTTCAAGAGCGCGG 762

## RESULT 8

US-09-902-540-2740  
Sequence 2740, Application US/09902540  
Patent No. 6833447

## GENERAL INFORMATION:

APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.

APPLICANT: Wiegand, Roger C.

TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof

FILE REFERENCE: 38-10(15849)B

CURRENT APPLICATION NUMBER: US/09/902,540

PRIOR FILING DATE: 2001-07-10

PRIOR APPLICATION NUMBER: 60/217,883

PRIOR FILING DATE: 2000-07-10

NUMBER OF SEQ ID NOS: 16825

SEQ ID NO 2740

LENGTH: 1963

TYPE: DNA

ORGANISM: Myxococcus xanthus

US-09-902-540-2740

Query Match 17.4%; Score 54.2; DB 4; Length 1963;  
Best Local Similarity 51.4%; Pred. No. 0.0039;  
Matches 125; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

Qy 1 ATGACCGAGACAGAGCTGACGCGGCTGCGCGGAGCTGACGCGCTTCGACGGGAGC 60  
Db 1 ATGACCGACCTTCCCGACCTTCGACCTTCGACGCGGAGCTTCGACGAGAG 60  
Qy 61 CTCCTGACACGAGTGGCGGCGGACCTGAGTGTGCGATGCGGCGGAGTACAGTCC 120  
Db 61 ATCTGTGATGCTTCCGCGCGGAGTGGCGGAGTGGCGGAGTGGCGGAGTGG 120  
Qy 121 CGGACGCGGCTTCCGATGATGACGCGCGGCGGCTGATCAAGAGACAGGCGCGC 180  
Db 121 GCGCGCGCGCGCTTCCGCGGAGCAGCGCGGAGAGTGTGCTGCGCGGAGTCCG 180  
Qy 181 CGCTAGCGCGCGGACGAGCGCTTCGAGCATGCTTCTGTGTAACCTTACGACGATC 240  
Db 181 ACCCGCGCGCGAGAGACGCGCTTCGACCGCGAGAGTGGAGGACATTCGCGGCTTTC 240  
Qy 241 ATC 243  
Db 241 ATC 243

## RESULT 9

US-09-902-540-1129/c  
Sequence 1129, Application US/09902540  
Patent No. 6833447

## GENERAL INFORMATION:

APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.

APPLICANT: Wiegand, Roger C.

TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof

FILE REFERENCE: 38-10(15849)B

Qy 1 ATGACCGAGACAGAGCTGACGCGGCTGCGCGGAGCTGACGCGCTTCGACGGGAGC 60  
Db 8044 ATGACCGACCTTCCCGACCTTCGACCTTCGACGCGGAGCTTCGACGAGAG 7985  
Qy 61 CTCCTGACACGAGTGGCGGCGGACCTGAGTGTGCGATGCGGCGGATACAGTCC 120  
Db 7984 ATCTGTGATGCTTCCGCGCGGAGTGGCGGAGTGGCGGAGTGGCGGAGTGG 7925  
Qy 121 CGGACGCGGCTTCCGATGATGACGCGCGGCGGCTGACGCTTGTGTAAGAGACGCGCGC 180  
Db 7924 GCGCGCGCGCGCTTCCGCGGAGCAGCGCGCGGAGAGTGTGCTTCGCGCGCATCCG 7865  
Qy 181 CGCTAGCGCGCGGACGAGCGCTTCGAGCATGCTTCTGTGTAACCTTACGACGATC 240  
Db 7864 ACCCGCGCGCGAGAGACGCGCTTCGAGCAGCGGAGTGGAGGACGATTCGCGGCTTTC 7805  
Qy 241 ATC 243  
Db 7804 ATC 7802

Query Match 17.4%; Score 54.2; DB 4; Length 15689;  
Best Local Similarity 51.4%; Pred. No. 0.0048;  
Matches 125; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

RESULT 10  
US-08-403-852D-4  
Sequence 4, Application US/08403852D  
Patent No. 5891695

## GENERAL INFORMATION:

APPLICANT: Blanc, Veronique  
APPLICANT: Blanche, Francis

APPLICANT: Cruzet, Joel  
APPLICANT: Jacques, Nathalie

APPLICANT: Lacroix, Patricia  
APPLICANT: Thibaut, Denis

APPLICANT: Zagorec, Montague  
APPLICANT: Debussche, Laurent

APPLICANT: De Crey-Lagard, Valerie

TITLE OF INVENTION: Polypeptides Involved In The

TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences

TITLE OF INVENTION: Coding For These Polypeptides And Their Use

NUMBER OF SEQUENCES: 43

CORRESPONDENCE ADDRESS:

ADDRESSER: Finnegan, Henderson, Farabow, Garrett & Dunner

STREET: 1300 I Street, N.W., Suite 700

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3315

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/403,852D  
FILING DATE: 10-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR 93/00923  
FILING DATE: 25-SEP-1993  
PRIOR APPLICATION DATA:



```
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1208 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1208

US-08-403-852D-4

Query Match      17.3% Score 54; DB 2; Length 1208;
Best Local Similarity 50.0%; Pred. No. 0.0042;
Matches 135; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 38 AGCTCGACGCCCTTCGACGGGACGCTCCTGACACGATGCGCGCCGCGATCGACCTCGGTG 97
    |||||
DB 860 AGGTGACCGCTTCGCGCGCGGACGATGGCGTGGTCCGCAAGAGTGGTCCCGCG 919
    |||||

QY 98 TCCGATCGCGCGGTACAGATCCCGGACGCGTCCGATGATGACCGCGCGGTCA 157
    |||||
DB 920 GCGTCGCTCCCGCTGCGAGGTCCAGGTCCGCTACGCGATCGCGACGCGCGGTG 979
    |||||

QY 158 GCGTGTCAAGGACGAGCGCGCGCTACGCGCGCGACGCGCGCTGCGAATCGTTC 217
    |||||
DB 980 GCGTGTTCGTCGACGCTTGGGACCGGACCGGCGCGCGATCGAGAGCGATCGAGAGCGCA 1039
    |||||

QY 218 TGGTGAACCTCTACGACGATGATCATCGAGAGATGTCGCGGTGAGAGACCTGGTATGA 277
    |||||
DB 1040 TCACCGAGGTCTTCGACCTGCGCGCGCGCGCGCATCATCGGACCTGCTGCGGC 1099
    |||||

QY 278 GCGGAGAGCGCTGACGCGCGGAGACCGGC 307
    |||||
DB 1100 CCATCTACGCCGCGCACCGCGCTACGCGC 1129
    |||||

RESULT 11
US-08-510-646B-4
; Sequence 4, Application US/08510646B
; Patent No. 6077699
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Blanche, Francis
; APPLICANT: Crouzet, Joel
; APPLICANT: Jacques, Nathalie
; APPLICANT: Lacroix, Patricia
; APPLICANT: Thibaut, Denis
; APPLICANT: Zagorec, Monique
; APPLICANT: Debussche, Laurent
; APPLICANT: De Crey-Lagard, Valerie
; TITLE OF INVENTION: Polypeptides Involved in The
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
; TITLE OF INVENTION: Coding For These Polypeptides And Their Use
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
```

```
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/510,646B
FILING DATE: 03-AUG-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,852
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1208 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1208

US-08-510-646B-4

Query Match      17.3% Score 54; DB 3; Length 1208;
Best Local Similarity 50.0%; Pred. No. 0.0042;
Matches 135; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 38 AGCTCGACGCCCTTCGACGGGACGCTCCTGACACGATGCGCGCGCGATCGACCTCGGTG 97
    |||||
DB 860 AGGTGACCGCTTCGCGCGCGGACGATGGCGTGGTCCGCAAGAGTGGTCCCGCG 919
    |||||

QY 98 TCCGATCGCGCGGTACAGTCCCGGACGCGCTCCGATGATGACCGCGCGGTCA 157
    |||||
DB 920 GCGTCGCTCCCGCTGCGAGGTCCAGGTGCGCTACGCGCATCGGACGCGCGGTG 979
    |||||

QY 158 GCGTGTCAAGGACGAGCGCGCGCTACGCGCGCGACGCGCGCTGCGAATCGTTC 217
    |||||
DB 980 GCGTGTTCGTCGACGCTTGGGACCGGACCGGCGCGCGATCGAGAGCGATCGAGAGCGCA 1039
    |||||

QY 218 TGGTGAACCTCTACGACGATGATCATCGAGAGATGTCGCGGTGAGAGACCTGGTATGA 277
    |||||
DB 1040 TCACCGAGGTCTTCGACCTGCGCGCGCGCGCGCATCATCGGACCTGCTGCGGC 1099
    |||||

QY 278 GCGGAGAGCGCTGACGCGCGGAGACCGGC 307
    |||||
DB 1100 CCATCTACGCCGCGCACCGCGCTACGCGC 1129
    |||||

RESULT 12
US-09-231-818-4
; Sequence 4, Application US/09231818
; Patent No. 6171846
; GENERAL INFORMATION:
```



APPLICANT: Blanc, Veronique  
APPLICANT: Blanche, Francis  
APPLICANT: Cruzet, Joel  
APPLICANT: Jacques, Nathalie  
APPLICANT: Lacroix, Patricia  
APPLICANT: Thibaut, Denis  
APPLICANT: Zagorec, Monique  
APPLICANT: Debussche, Laurent  
APPLICANT: De Crecy-Lagard, Valerie  
TITLE OF INVENTION: Polypeptides Involved In The  
Biosynthesis Of Streptogramins, Nucleotide Sequences  
Coding For These Polypeptides And Their Use  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/231,818  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/403,852  
FILING DATE: 10-MAY-1995  
APPLICATION NUMBER: PCT/FR 93/00923  
FILING DATE: 25-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/11441  
FILING DATE: 25-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03806.0054-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1208 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: S.pristinaespiralis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1208  
US-09-231-818-4

Query Match 17.3%; Score 54; DB 3; Length 1208;  
Best Local Similarity 50.0%; Pred. No. 0.0042;  
Matches 135; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 38 AGCTCGACGGCCTTCGACGGAGCGTCTCTGACACGGGCGGGCCCGCATCGACCTGGTG 97  
DB 860 AGGTCAGACCTTCCTCCGCCGCTACCGCATGCGCTGGTCCGCAAGAGTGTGCGCGG 919  
QY 98 TCGCATTCGGCGCGTACCAATCCCGGACGCGCTCCCGATGATGACAGCCCGCGGATCA 157  
DB 920 GCTCGGCTCCGCTGAGAGTCCAGGTGCTTACGCGCATCGGCAAGCGGACCGGTCG 979  
QY 158 GCTGTCAAGGACAGGGCGCGCCGCTACCGCGCCGACGCGGCTTCGACGAATCGTTC 217

DB 980 GCCTGTTCGAGAGCTTGGGACCGGACCGGTCCGCCAGAGCCCATCGAAGGCCA 1039  
QY 218 TGGTAACCTCTACGACGTGATCATCAGAGATGTCCGCGTCGAGACCTGTGATGA 277  
DB 1040 TCACGAGGTCTTCGACCTGCGCGCCCGGCCATATCCGCACTTCGACTGTGCGGC 1099  
QY 278 GCGGAGAGCTTCGACGCGCGGACCGGC 307  
DB 1100 CCATCTACCGCGCACCGCGCGCTACGCGCC 1129

RESULT 13  
US-09-635-359B-4  
Sequence 4, Application US/09635359B  
Patent No. 6670157  
GENERAL INFORMATION:  
APPLICANT: Blanc, Veronique  
Blanche, Francis  
Cruzet, Joel  
Jacques, Nathalie  
Lacroix, Patricia  
Thibaut, Denis  
Zagorec, Monique  
Debussche, Laurent  
De Crecy-Lagard, Valerie  
TITLE OF INVENTION: Polypeptides Involved In The  
Biosynthesis Of Streptogramins, Nucleotide Sequences  
Coding For These Polypeptides And Their Use  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/635,359B  
FILING DATE: 09-Aug-2000  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/231,818  
FILING DATE: 15-JAN-1999  
APPLICATION NUMBER: US 08/403,852  
FILING DATE: 10-MAY-1995  
APPLICATION NUMBER: PCT/FR 93/00923  
FILING DATE: 25-SEP-1993  
APPLICATION NUMBER: FR 92/11441  
FILING DATE: 25-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03806.0054-03000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1208 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: S.pristinaespiralis  
FEATURE:  
NAME/KEY: CDS



LOCATION: 1..1208  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-635-359B-4

Query Match 17.3%; Score 54; DB 4; Length 1208;  
Best Local Similarity 50.0%; Pred. No. 0.0042;  
Matches 135; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 38 AGCTGACGCGCTTGACGAGGACGCTCTGACACAGGTGCGGCGCCGATGACCTCGGTG 97  
DB 860 AGGTGACCGCTTCCGCGGACGATGCGCTGCGTCCAGAGACGTCCTCCCGCG 919  
QY 98 TCCGATCGCGCGGTACAGTCCCGGACGCGGTCCGATGATGACGCGCGCGGTCA 157  
DB 920 GCTTCGCTCCCGCTGCGAGGTCCAGGTGCGCTTCCGATGCGAGGCGCGGTG 979  
QY 158 GCGTGTCAAGAGAGGCGCGCGCTACGCGCGGACGAGCGCTGACGATCGTTC 217  
DB 980 GCGTGTGTCGACGAGTTCGCGACCGGACGTCGCGCGAGAGCGATGAGAAAGCCA 1039  
QY 218 TGGTGAACCTTACGAGTATCATACGAGATGTGCGCGGTGAGGACTGTGATGA 277  
DB 1040 TCACCGAGGTCTTGTGACCTGCGCGCGGACATCATCCGACCTGCTGCGGC 1099  
QY 278 GCCGGAGAGCGCTGACGCGCGAGGACCGGC 307  
DB 1100 CCATCTACGCGCGCCACCGCGCTACGCGC 1129

RESULT 14  
US-08-403-852D-1  
; Sequence 1, Application US/08403852D  
; Patent No. 5891695  
; GENERAL INFORMATION:  
; APPLICANT: Blanc, Veronique  
; APPLICANT: Blanche, Francis  
; APPLICANT: Crouzet, Joel  
; APPLICANT: Jacques, Nathalie  
; APPLICANT: Lacroix, Patricia  
; APPLICANT: Thibaut, Denis  
; APPLICANT: Zagorec, Monique  
; APPLICANT: Debussche, Laurent  
; APPLICANT: De Crecy-Lagard, Valerie  
; TITLE OF INVENTION: Polypeptides Involved In The  
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
; NUMBER OF INVENTION: Coding For These Polypeptides And Their Use  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/403,852D  
; FILING DATE: 10-MAY-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/FR 93/00923  
; FILING DATE: 25-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 92/11441  
; FILING DATE: 25-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meyers, Kenneth J.  
; REGISTRATION NUMBER: 25,146  
; REFERENCE/DOCKET NUMBER: 03806.0054-00000  
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5392 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: S.pristinaespiralis  
US-08-403-852D-1

Query Match 17.3%; Score 54; DB 2; Length 5392;  
Best Local Similarity 50.0%; Pred. No. 0.0048;  
Matches 135; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 38 AGCTGACGCGCTTGACGAGGACGCTCTGACACGGTGCGGCGCCGATGACCTCGGTG 97  
DB 4417 AGGTGACCGCTTCCGCGGACGATGCGCTGCGTCCGACGAGACGTCGTGCGCGG 4476  
QY 98 TCCGATCGCGCGGTACAGTCCCGGACGCGGTCCGATGATGACGCGCGCGGTCA 157  
DB 4477 GCTTCGCTCCCGCTGCGAGGTCCAGGTGCGCTTACGCTACGCGGACGCGGTG 4536  
QY 158 GCGTGTCAAGAGAGGCGCGCGCTACGCGCGGACGAGCGCTGACGATCGTTC 217  
DB 4537 GCGTGTGTCGAGAGGTTCGCGACCGGACGCGTCCGACGAGCGGATGAGAAAGCCA 4596  
QY 218 TGGTGAACCTTACGAGTATCATACGAGATGTGCGCGGTGAGGACTGTGATGA 277  
DB 4597 TCACCGAGGTCTTGTGACCTGCGCGCGGACATCATCCGACCTGACCTGCTGCGGC 4656  
QY 278 GCCGGAGAGCGCTGACGCGCGAGGACCGGC 307  
DB 4657 CCATCTACGCGCGCCACCGCGCTACGCGC 4686

RESULT 15  
US-08-510-646B-1  
; Sequence 1, Application US/08510646B  
; Patent No. 6077699  
; GENERAL INFORMATION:  
; APPLICANT: Blanc, Veronique  
; APPLICANT: Blanche, Francis  
; APPLICANT: Crouzet, Joel  
; APPLICANT: Jacques, Nathalie  
; APPLICANT: Lacroix, Patricia  
; APPLICANT: Thibaut, Denis  
; APPLICANT: Zagorec, Monique  
; APPLICANT: Debussche, Laurent  
; APPLICANT: De Crecy-Lagard, Valerie  
; TITLE OF INVENTION: Polypeptides Involved In The  
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
; NUMBER OF INVENTION: Coding For These Polypeptides And Their Use  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/510,646B  
; FILING DATE: 03-AUG-1995



CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/403,852  
FILING DATE: 10-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR 93/00923  
FILING DATE: 25-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/11441  
FILING DATE: 25-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03806.0054-01000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
FAX: (202) 408-4000  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5392 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: S.pristinaeapiralis  
US-08-510-646B-1

Query Match 17.3%; Score 54; DB 3; Length 5392;  
Best Local Similarity 50.0%; Pred. No. 0.0048;  
Matches 135; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 38 AGCTTCAGCGCCCTCGACGGGACGCTCTGTGACACAGTGGCGGCGCCGATGACCTGGTG 97  
DB 4417 AGGTTCAGCGCCCTCGACGGGACGCTCTGTGACACAGTGGCGGCGCCGATGACCTGGTG 4476  
QY 98 TCCGATTCGCGCGGTACAGTCCCGGACCGCGTCCCGATGATGACAGCCCGCGGAGTCA 157  
DB 4477 GCTTCGCGTCCCGGTGAGAGTCCAGTCCCTACCGCATCGGCGAGCCGAGCGGTG 4536  
QY 158 GCTGTGTCAGAGACAGGCGCGCGCTACGCGCGCGACACAGGCTTCGACGAATCGTTCC 217  
DB 4537 GCTGTGTCGTCAGAGCGTTCGACACCGGACCGCGCGCGAGCGCATCGAGAGGCCA 4596  
QY 218 TGGTGAACTCTACAGAGTATCATACAGAGATGTCCCGCTCGAGAGCTGTGATGA 277  
DB 4597 TCAACGAGTCTTCGACCTGCGCGCGCGCATCATCTGACCTGACCTGTGCGGC 4656  
QY 278 GCCGAGAGAGCTGACCGCGCGAGAGACCGGC 307  
DB 4657 CCATCTACGCGCGCGCGCGCGCTACGCGC 4686

Search completed: October 5, 2005, 10:06:41  
Job time : 59.8815 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 20:50:02 : Search time 306.305 Seconds  
(without alignments)  
7033.027 Million cell updates/sec

Title: US-10-089-514-3

Perfect score: 312  
Sequence: 1 atgacgcgagcagcagcagc.....cggcgcgagcgcgcgcgtga 312

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7442561 seqs, 3452328358 residue

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/prodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/prodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/prodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/prodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/prodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/prodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/prodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/prodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
15: /cgn2\_6/prodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
16: /cgn2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/prodata/1/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgn2\_6/prodata/1/pubpna/US10F\_PUBCOMB.seq:\*  
19: /cgn2\_6/prodata/1/pubpna/US10G\_PUBCOMB.seq:\*  
20: /cgn2\_6/prodata/1/pubpna/US10H\_PUBCOMB.seq:\*  
21: /cgn2\_6/prodata/1/pubpna/US10I\_PUBCOMB.seq:\*  
22: /cgn2\_6/prodata/1/pubpna/US10I\_NEW\_PUB.seq:\*  
23: /cgn2\_6/prodata/1/pubpna/US11\_PUBCOMB.seq:\*  
24: /cgn2\_6/prodata/1/pubpna/US11A\_PUBCOMB.seq:\*  
25: /cgn2\_6/prodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
26: /cgn2\_6/prodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	312	100.0	312	US-10-472-587-3	Sequence 3, Appl1
2	296.4	95.0	3305	US-10-126-927-68	Sequence 68, Appl1
3	284.8	84.9	12391	US-10-126-927-67	Sequence 67, Appl1
4	59.4	19.0	3114	US-10-437-963-48039	Sequence 48039, A
5	59.2	19.0	838	US-10-425-115-320	Sequence 320, App
6	57.6	18.5	1009	US-10-425-114-33332	Sequence 33332, A
7	55	17.6	1398	US-10-156-761-1302	Sequence 1302, Ap

Result No.	Score	Query Match	Length	ID	Description
8	55	17.6	9025608	US-10-156-761-1	Sequence 1, Appl1
9	54.8	17.6	1458	US-10-107-431-76	Sequence 76, Appl1
10	54.8	17.6	45055	US-10-107-431-277	Sequence 277, Appl1
11	54.8	17.6	9025608	US-10-156-761-1	Sequence 1, Appl1
12	54.4	17.4	1284	US-10-156-761-3832	Sequence 3832, Appl1
13	54	17.3	1208	US-10-716-803-4	Sequence 4, Appl1
14	54	17.3	5392	US-10-716-803-1	Sequence 1, Appl1
15	53.6	17.2	837	US-10-107-431-112	Sequence 112, Appl1
16	53.6	17.2	2817	US-10-156-761-4836	Sequence 4836, Ap
17	53.6	17.2	29870	US-10-494-495-1	Sequence 1, Appl1
18	53.6	17.2	45055	US-10-107-431-277	Sequence 277, Appl1
19	53	17.0	1356	US-10-282-122A-30160	Sequence 30160, A
20	52.6	16.9	1869	US-10-369-493-28376	Sequence 28376, A
21	52.6	16.9	1869	US-10-369-493-31135	Sequence 31135, A
22	52.6	16.9	2264	US-10-437-963-30078	Sequence 30078, A
23	52.6	16.9	3735	US-10-369-493-44134	Sequence 44134, A
24	52.4	16.8	1410	US-10-680-860A-40	Sequence 40, Appl1
25	52.4	16.8	30943	US-10-680-860A-1	Sequence 1, Appl1
26	52.2	16.7	762	US-10-156-761-3402	Sequence 3402, Ap
27	51.8	16.6	1686	US-10-494-495-4	Sequence 4, Appl1
28	51.6	16.5	732	US-10-156-761-1775	Sequence 1775, Ap
29	51.6	16.5	13654	US-10-292-798-1131	Sequence 1131, Ap
30	51.4	16.5	1393	US-10-437-963-8368	Sequence 8368, Ap
31	51.2	16.4	1707	US-10-156-761-4723	Sequence 4723, Ap
32	51.2	16.4	5562	US-09-030-482B-18	Sequence 18, Appl1
33	51.2	16.4	6816	US-09-935-541-1	Sequence 1, Appl1
34	51.2	16.4	6816	US-10-425-800-1	Sequence 1, Appl1
35	51.2	16.4	6855	US-09-935-541-3	Sequence 3, Appl1
36	51.2	16.4	6855	US-10-425-800-3	Sequence 3, Appl1
37	51.2	16.4	6990	US-10-377-139-8	Sequence 8, Appl1
38	50.8	16.3	862	US-10-156-761-6124	Sequence 6124, Ap
39	50.8	16.3	3012	US-10-156-761-2716	Sequence 2716, Ap
40	50.8	16.3	6122	US-10-225-567A-179	Sequence 179, Ap
41	50.8	16.3	6122	US-10-755-889-185	Sequence 185, App
42	50.8	16.3	6122	US-10-722-357-23	Sequence 23, Appl1
43	50.6	16.2	1014	US-10-156-761-3014	Sequence 3014, Ap
44	50.6	16.2	1414	US-10-739-930-2200	Sequence 2200, Ap
45	50.6	16.2	1537	US-10-425-115-110407	Sequence 110407,

#### ALIGNMENTS

RESULT 1  
US-10-472-587-3  
; Sequence 3, Application US/10472587  
; Publication No. US20040214274A1  
; GENERAL INFORMATION:  
; APPLICANT: YANAI, Koji  
; APPLICANT: SUMIDA, Naomi  
; APPLICANT: MATANABE, Manabu  
; APPLICANT: MORIYA, Tatsuaki  
; APPLICANT: MORAKAMI, Takehshi  
; TITLE OF INVENTION: Transformants Producing Substance P/1022 Derivatives, Methods fo  
; FILE REFERENCE: 2003-1302A/MMC/00144  
; CURRENT FILING DATE: 2003-09-22  
; PRIOR APPLICATION NUMBER: 82227/2001  
; PRIOR FILING DATE: 2001-03-22  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 312  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(309)  
US-10-472-587-3  
Query Match 100.0%; Score 312; DB 20; Length 312;  
Best Local Similarity 100.0%; Pred. No. 5.1e-75;



Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGACCGAGCAAGACGAGCTGACGCGCTGCGCGAGCTCGACGCCCTCGACGGAGCG 60
Db 1 ATGACCGAGCAAGACGAGCTGACGCGCTGCGCGAGCTCGACGCCCTCGACGGAGCG 60
QY 61 CTCTTGACACGGTGGCGCGCGCGAGCTGACCTCGGTGTCCGCAATCGCGCGGTACAAGTCC 120
Db 61 CTCTTGACACGGTGGCGCGCGCGAGCTGACCTCGGTGTCCGCAATCGCGCGGTACAAGTCC 120
QY 121 CGGACCGCGCTCCCGATGATGACGCGCGCGGTGACGCTGCTCAAGGACAGGCGCGCC 180
Db 121 CGGACCGCGCTCCCGATGATGACGCGCGCGGTGACGCTGCTCAAGGACAGGCGCGCC 180
QY 181 CGCTACGCGCGCGACCGAGCTCGACGATCGTTCTGTGTGAACCTCTACGACGTGATC 240
Db 181 CGCTACGCGCGCGACCGAGCTCGACGATCGTTCTGTGTGAACCTCTACGACGTGATC 240
QY 241 ATGACCGAGATGTGCGCGCTGACGAGACCTGTGTATGACCGGAGAGCCTTGACGGCGAG 300
Db 241 ATGACCGAGATGTGCGCGCTGACGAGACCTGTGTATGACCGGAGAGCCTTGACGGCGAG 300
QY 301 GACCGCGCGTGA 312
Db 301 GACCGCGCGTGA 312
```

## RESULT 2

```
US-10-126-927-68
; Sequence 68, Application US/10126927
; Publication No. US2003082575A1
; GENERAL INFORMATION:
; APPLICANT: The Scripps Research Institute
; APPLICANT: Schultz, Peter G
; APPLICANT: Wang, Lei
; APPLICANT: Anderson, John C
; APPLICANT: Chin, Jason
; APPLICANT: Liu, David R
; APPLICANT: Magliery, Thomas
; APPLICANT: Meggers, Eric L
; APPLICANT: Mehl, Ryan A
; APPLICANT: Pasternak, Miro
; APPLICANT: Santoro, Stephen W
; APPLICANT: Zhang, Zhiwen
; TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
; FILE REFERENCE: 54-000120US
; CURRENT APPLICATION NUMBER: US/10/126,927
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,030
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 60/355,514
; PRIOR FILING DATE: 2002-02-06
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 68
; LENGTH: 3305
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-10-126-927-68
```

Query Match 95.0%; Score 296.4; DB 14; Length 3305;  
Best Local Similarity 98.4%; Pred. No. 7.6e-71;  
Matches 311; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

```
QY 1 ATGACCGAGCAAGACGAGCTGCA-----GCGGCTGCGCGCGAGACTTGACGCCCTCGACGG 56
Db 2990 ATGACCGAGCAAGACGAGCTGCAAGTTGCGGCTGCGCGAGACTTGACGCCCTCGACGG 3049
QY 57 GACGCTCTTGACACGAGTGGCGCGCGCATGACCTCGGTGTCCGCAATCGCGCGGTACAA 116
Db 3050 GACGCTCTTGACACGAGTGGCGCGCGCATGACCTCGGTGTCCGCAATCGCGCGGTACAA 3109
QY 117 GTCCCGGACGCGGTCCGATGATGACAGCCCGGCGGCTGAGCTGTCAAGAGACAGGCG 176
```

```
Db 3110 GTCCCGGACGCGGTCCGATGATGACAGCCCGCGGCTGACCTGTGTCAAGAGACAGGCG 3169
QY 177 CGCCGCTGACGCGCGCGACCAAGGCTCGACGAAATGTTCTGTGTGAACCTCTACAGCGT 236
Db 3170 CGCCGCTGACGCGCGCGACCAAGGCTCGACGAAATGTTCTGTGTGAACCTCTACAGCGT 3229
QY 237 GATCATCAAGAGATGTGCGCGGTGAGAGACCTGTGTATGAGCCGAGAGCCTGACGCGC 296
Db 3230 GATCATCAAGAGATGTGCGCGGTGAGAGACCTGTGTATGAGCCGAGAGCCTGACGCGC 3289
QY 297 CGAGGACCGCGCGTGA 312
Db 3290 CGAGGACCGCGCGTGA 3305
```

## RESULT 3

```
US-10-126-927-67
; Sequence 67, Application US/10126927
; Publication No. US2003082575A1
; GENERAL INFORMATION:
; APPLICANT: The Scripps Research Institute
; APPLICANT: Schultz, Peter G
; APPLICANT: Wang, Lei
; APPLICANT: Anderson, John C
; APPLICANT: Chin, Jason
; APPLICANT: Liu, David R
; APPLICANT: Magliery, Thomas
; APPLICANT: Meggers, Eric L
; APPLICANT: Mehl, Ryan A
; APPLICANT: Pasternak, Miro
; APPLICANT: Santoro, Stephen W
; APPLICANT: Zhang, Zhiwen
; TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
; FILE REFERENCE: 54-000120US
; CURRENT APPLICATION NUMBER: US/10/126,927
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,030
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 60/355,514
; PRIOR FILING DATE: 2002-02-06
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67
; LENGTH: 12391
; TYPE: DNA
; ORGANISM: Plasmid pSCI01, Streptomyces venezuelae papABC
US-10-126-927-67
```

Query Match 84.9%; Score 264.8; DB 14; Length 12391;  
Best Local Similarity 97.9%; Pred. No. 2.6e-62;  
Matches 280; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

```
QY 1 ATGACCGAGCAAGACGAGCTGCA---GCGGCTGCGCGCGAGCTTGACGCCCTCGACGG 56
Db 2055 ATGACCGAGCAAGACGAGCTGCAAGTTGCGGCTGCGCGAGCTTGACGCCCTCGACGG 2114
QY 57 GACGCTCTTGACACGAGTGGCGCGCGCATGACCTCGGTGTCCGCAATCGCGCGGTACAA 116
Db 2115 GACGCTCTTGACACGAGTGGCGCGCGCATGACCTCGGTGTCCGCAATCGCGCGGTACAA 2174
QY 117 GTCCCGGACGCGGTCCGATGATGACAGCCCGGCGGTGACGCTGTGTCAAGAGAGGCG 176
Db 2175 GTCCCGGACGCGGTCCGATGATGACAGCCCGGCGGTGACGCTGTGTCAAGAGAGGCG 2234
QY 177 CGCCGCTGACGCGCGCGACCAAGGCTCGACGAAATGTTCTGTGTGAACCTCTACAGCGT 236
Db 2235 CGCCGCTGACGCGCGCGACCAAGGCTCGACGAAATGTTCTGTGTGAACCTCTACAGCGT 2294
QY 237 GATCATCAAGAGATGTGCGCGGTGAGAGACCTGTGTATGAGCCG 282
Db 2295 GATCATCAAGAGATGTGCGCGGTGAGAGACCTGTGTATGAGCCG 2340
```



## RESULT 4

US-10-437-963-48039  
; Sequence 48039, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 48039  
; LENGTH: 3114  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_50752C.1  
US-10-437-963-48039

Query Match 19.0%; Score 59.4; DB 19; Length 3114;  
Best Local Similarity 50.2%; Pred. No. 1.1e-06;  
Matches 147; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 16 GAGCTGACGCGGCTCCGCGGAGCTGACGCGCTCCGAGGAGCGCTCTGGAACGCTG 75  
DB 853 GAGGAG 912  
QY 76 CGGCGCGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 135  
DB 913 GAGTTACGCGGAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 972  
QY 136 ATGATGACGCGGCGGCGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 195  
DB 973 GTCAAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1032  
QY 196 CACGCGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 255  
DB 1033 GTCGCGCTGCGCGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1092  
QY 256 CGGCTGAGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 308  
DB 1093 GGCCTGAGTGAAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1145

## RESULT 5

US-10-425-115-320  
; Sequence 320, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 320  
; LENGTH: 838  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:

OTHER INFORMATION: Clone ID: MRT4577\_100296C.1  
US-10-425-115-320

Query Match 19.0%; Score 59.2; DB 20; Length 838;  
Best Local Similarity 53.7%; Pred. No. 1.4e-06;  
Matches 146; Conservative 0; Mismatches 123; Indels 3; Gaps 1;

QY 34 GCGGAGCTGAGCGCTCCGAGGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 93  
DB 313 GTGAGAGAGCTGCGCGCGGAGAGTCTGCTATGCGATGAGTGTGAGTGTGAGTGTGAGTGTG 372  
QY 94 GGTGTCGAGCTGCGCGGAGTCAAGTCCGAGCGAGCTGCGAGTGTGAGTGTGAGTGTGAGTGTG 153  
DB 373 GCGATCCGAGAGGTGTGCGCGAGCGCGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 432  
QY 154 GTACGCTGTGAGAGACAGAGCGCGCGCTGAGC--GCGGACACAGCGCTGAGAG 210  
DB 433 GCGGCGAGCTCTTCCGCAAGGTGCGCGCGCGCTGAGCTGAGAGACCGCGCGCGAG 492  
QY 211 TCGTTCCTGAGTCAAGCTTCAAGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATG 270  
DB 493 TACTTCAGAGCTGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 552  
QY 271 GTGATGAGCGCGAGAGAGCTGAGCGCGAGGA 302  
DB 553 GACCTCAAGCGCGAGAGAGCTGCTGCTGAGGA 584

## RESULT 6

US-10-425-114-33332  
; Sequence 33332, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaka, Jack E  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 33332  
; LENGTH: 1009  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17095G02\_FLI1  
US-10-425-114-33332

Query Match 18.5%; Score 57.6; DB 18; Length 1009;  
Best Local Similarity 53.6%; Pred. No. 3.7e-06;  
Matches 120; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 83 GCATGACCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTG 142  
DB 537 GCAACGAGCTGCGCGGAGCTGAGTGTGAGGCGAGTGTGAGCGCGAGTGTGAGTGTGAGTGTG 596  
QY 143 AGCGCGCGCGGCTGAGCTGAGTGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 202  
DB 597 TCCGCGAGCGCGCGCGCGCGAGAGATGAGAGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTG 656  
QY 203 TCGAGAGTGTGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 262  
DB 657 CCGAGCTTCTTCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 716  
QY 263 AGGAGCTGTGATGAGCGCGAGAGAGCTGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 306  
DB 717 TGGCGCTGGGGTTCAGTGTGAGCGCACTCTCATGCGGAGCATCCCG 760



```
RESULT 7
US-10-156-761-1302
; Sequence 1302, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1302
; LENGTH: 1398
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1398)
US-10-156-761-1302
```

```
Query Match          17.6%; Score 55; DB 15; Length 1398;
Best Local Similarity 48.8%; Pred. No. 1.8e-05;
Matches 148; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
```

```
QY 9 GCAGAACGAGCTGCAGCGGCTGCGCGGAGCTGCAGCCCTGCAGCGGACGCTCTCTGA 68
    |||||
DB 423 GCAGACACCCGCCACCCGCTGACGACCGGCTCTCTGCGCGCTGCGCCCTGTCGCG 482
    |||||
QY 69 CACGGTGGCGCGCGCGCATGCACTCTGCGCATGCGCGGATCAAGTCCCGGACGCG 128
    |||||
DB 483 CTTGACCGCGCGGAGCTGCGCGCGCGGCTGCTGCGCATGCGCTGCTGCTGCG 542
    |||||
QY 129 CGTCCGATGATGACGCGCGCGCGGCTGCACTGTCAGAGAGAGGCGCGCGCTAAC 188
    |||||
DB 543 GGTGGGCTGATGCGCGCGCGGACATGCCCGGCTGCGCGGACAGAGCGCGCTGAC 602
    |||||
QY 189 CGCGGACGAGCGCTGCAGCAATGCTTCTGTGAACCTTACGACGTGATCAACGA 248
    |||||
DB 603 CGCGCGCTGTGCTGCGCGCTCTCCACCCGCTGTGCTGCTGCTGCTGCTGCTG 662
    |||||
QY 249 GATGTGCGCGGTGAGAGACCTGTGATGAGCGCGGAGAGCTGACGCGCGAGAGCC 308
    |||||
DB 663 GCACAAAGACCGCGTCACTGCTGGGGCTGCTGGGCGCGGACTGATGCGCGCGCG 722
    |||||
QY 309 GTG 311
    |||
DB 723 GTG 725
```

```
RESULT 8
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
```

```
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
```

```
Query Match          17.6%; Score 55; DB 15; Length 9025608;
Best Local Similarity 48.8%; Pred. No. 1.2e-05;
Matches 148; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
```

```
QY 9 GCAGAACGAGCTGCAGCGGCTGCGCGGAGCTGCAGCCCTGCAGCGGACGCTCTCTGA 68
    |||||
DB 1624189 GCAGCACAGCGCGACCCCGCTGACGACCGGCTCTCTGCGCGCTGCTGTCGG 1624130
    |||||
QY 69 CACGGTGGCGCGCGCGCATGACCTGCGATGCGCGGATCAAGTCCCGGACGCG 128
    |||||
DB 1624129 CTTGACCGCGCGGAGCTGCGCGCGGCTGCTGCGCATGCGGCTGCTGCTGCG 1624070
    |||||
QY 129 CGTCCGATGATGACGCGCGCGGCTGCACTGTCAGAGAGAGAGGCGCGCGCTAAC 188
    |||||
DB 1624069 GGTGGGCTGATGAGGCGCGGACCTGCGCGCGGCTGCGCGGACAGAGGCGTGCAG 1624010
    |||||
QY 189 CGCGGACGAGCGCTGCAGCAATGCTTCTGTGAACCTTACGACGTGATCAACGA 248
    |||||
DB 1624009 CGCGGCTGTGATGCTGCGCGCTCTCCACCCGCTGTGCTGCTGCTGCTGCTG 1623950
    |||||
QY 249 GATGTGCGCGGTGAGAGACCTGTGATGAGCGGAGAGCTGACGCGCGAGAGCCGCG 308
    |||||
DB 1623949 GCACAAAGACCGCGTCACTGCTGGGGCTGCTGGGCGCGGACTGATGCGCGCG 1623890
    |||||
QY 309 GTG 311
    |||
DB 1623889 GTG 1623887
```

```
RESULT 9
US-10-107-431-76
; Sequence 76, Application US/10107431
; Publication No. US20030224364A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO
; FILE REFERENCE: 3001-7US
; CURRENT APPLICATION NUMBER: US/10/107,431
; CURRENT FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 76
; LENGTH: 1458
; TYPE: DNA
; ORGANISM: Streptomyces mobaraensis
US-10-107-431-76
```

```
Query Match          17.6%; Score 54.8; DB 17; Length 1458;
Best Local Similarity 49.0%; Pred. No. 2.1e-05;
Matches 146; Conservative 0; Mismatches 152; Indels 0; Gaps 0;
```

QY 6 CGAGGAGACGAGCTGCAGCGGCTGCGCGGAGCTGCAAGCCCTGCAGCGGACGCTCTCT 65



```
Db      684  CCACGAGACCGCTTCTTGAGCGCCCGAAGAGACAGACGACTGGAGGGGATGCTCGC 743
        66  GGACACGGGTGGGGCGCCGATCGACTCTGGTGTCCGATCGCGGGTACAAAGTCCCGCA 125
        Db      744  CTCGCTGCTCCAGAGAACCCGACGCTCCGACCTCATGGCGGAGGCGATACAGCC 803
        Qy      126  CGGCGTCCCGATGATGAGCCCGGCGGGTCAAGCTGATCAAGACAGGGCGCCGCTA 185
        Db      804  GGGGCTCCGGGTGAGAGAGACTTCTCTACGTGCGGACAGCTTCCACGGCCCGGCTA 863
        Qy      186  CGCGCGGACGACGCGCTCGACGAATGTTCTGATGAACCTCTACGACGTGATCATCAC 245
        Db      864  CTACCTGCGCGGACGCGCGCTCTCTCGACCCGCGTGTGTCACCGGCGTCCACT 923
        Qy      246  GGAAGATGTGCGCGCTCGACGACCTGTGTATGAGCCGGAGAGCCTGACGCGCCGAGAC 303
        Db      924  GGGCGCTCTACAGCGGACATGTGCGCGGACGCTCGTGTCTCGCACGGTGAACGAGAG 981
```

## RESULT 10

```
US-10-107-431-277/c
; Sequence 277, Application US/10107431
; Publication No. US2003024364A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Farnet, Chris
; APPLICANT: Staefle, Alfred
; APPLICANT: Zaopolou, Emmanuel
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO
; TITLE OF INVENTION: BIOSYNTHETIC LOCI
; FILE REFERENCE: 3001-7US
; CURRENT APPLICATION NUMBER: US/10/107,431
; CURRENT FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 277
; LENGTH: 45055
; TYPE: DNA
; ORGANISM: Streptomyces mobaraensis
US-10-107-431-277
```

```
Query Match      17.6%; Score 54.8; DB 17; Length 45055;
Best Local Similarity 49.0%; Pred. No. 1.8e-05;
Matches 146; Conservative 0; Mismatches 152; Indels 0; Gaps 0;
```

```
Qy      6  CGACGACGACGAGCTGACGCGCTGCGCGGCGAGCTGACGCGCTCGACGCGACGCTCT 65
Db      37629  CCACGAGACCGCTTCTCTGAGCGCCGCAAGAGACAGACGACCTGAGGCGATGCTCGC 37570
Qy      66  GGAACGCGTGGCGCGCGGACCTCGACCTCGGTGTCGATCGCGCGGTACAAAGTCCCGCA 125
Db      37569  CTCGCTGCTCCAGAGAACCCGACCGTCCGACCTCATGGCGGAGGCGGACGATACGACC 37510
Qy      126  CGGCGTCCCGATGATGAGCCCGGCGGGTCAAGCTGATCAAGACAGGGCGCCGCTA 185
Db      37509  GGGGCTCCGGGTGAGAGAGACTTCTCTACGTGCGGACAGCTTCCACGGCCCGGCTA 37450
Qy      186  CGCGCGGACGACGCGCTCGACGAATGTTCTGATGAACCTCTACGACGTGATCATCAC 245
Db      37449  CTACCTGCGCGGACGCGCGCTCTCTCGACCCGCGTGTGTCACCGGCGTCCACT 37390
Qy      246  GGAAGATGTGCGCGCTCGACGACCTGTGTATGAGCCGGAGAGCCTGACGCGCCGAGAC 303
Db      37389  GGGCGCTCTACAGCGGACATGTGCGCGGACGCTCGTGTCTCGCACGGTGAACGAGAGAC 37332
```

## RESULT 11

```
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
```

```
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
```

```
Query Match      17.6%; Score 54.8; DB 15; Length 9025608;
Best Local Similarity 50.0%; Pred. No. 1.4e-05;
Matches 137; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
```

```
Qy      33  CGGAGACTCGACCGCTCGACCGGACGCTCTGACACAGTGGCGCGCGCATCGACCT 92
Db      4753410  CACGCTGCTCAAGAGAGTGTGAGTGGCTCGACCGCGAGGTGAGCGCCGTACCGTCCG 4753469
Qy      93  CGGTGTCGATGCGCGCGGTACAAATGCCGACGCGCTGCCGATATGACGCCGCGCG 152
Db      4753470  GCGGCGCTGAGAGAGCGGTCAAGGTGAGAGACAGCGCGCTTCAACGAGGCTCAGCTG 4753529
Qy      153  GGTACGCTGTGTAAGAGACAGGCGCGCGCTACGCGCGCGACCAAGGCTCGACGAAATC 212
Db      4753530  GCGCTTGAAGACATACAGCGCGGTGCTGCACTTCCCGCGTGTGTAACGCCGACGCTCAC 4753589
Qy      213  GTTCTGTGTAACCTTACGACGTGATCATACAGAGATGTGCGGCTGAGAGACCTGCT 272
Db      4753590  CTTCCAGGTGTGTAAGGAGGCGCGCTGTGAGCGCGCGCGGCGGCTTGTGATGTGTC 4753649
Qy      273  GATGAGCGCGGAGAGCTTACGCGCGGACCGG 306
Db      4753650  GAAAGTCTTGAAGAGCAAGCGTACTGAGGCCCG 4753683
```

## RESULT 12

```
US-10-156-761-3832
; Sequence 3832, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3832
; LENGTH: 1284
; TYPE: DNA
```



ORGANISM: Streptomyces avermectilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1) .. (1284)  
US-10-156-761-3832

Query Match 17.4%; Score 54.4; DB 15; Length 1284;  
Best Local Similarity 50.8%; Pred. No. 2,7e-05;  
Matches 130; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

33 CGGGAGCTCGACGCGCCCTCGACGGGAGCTCTTGAGACATCGCGCGGCATCGACT 92  
1020 CACGTCCTCAAGAGGTCGTCGATCGTCGACGGGAGATAGCGCCGCTACCGTCG 1079

93 CGGTGTCGCGATCGCGCGGTACCAAGTCCCGSCACAGCGGTCCGATGATGACCCGCG 152  
1080 GCGGCGCTTGAGACGCGCGCTCAAGGTGAGACAGGCGGCTTCAACCGAGGCTCACTG 1139

153 GGTACGCTGTGTCAGACAGGCGCGCGCTACGCGCGGACCGACCGCTCGACGATC 212  
1140 GCGCTTCGAGGACATGACGCGGCTGTCGACTTCCCGCGTGTGTCACCGCGATCAG 1199

213 GTTCCTGTGATCTCTACGACGATGATCATCAGGAGATGTGCGCGCTCGAGACCTGCT 272  
1200 CTTCCAGTCTGTACGCGAGGCGCGCTGTGTGGCGCGCGCGCTTCTGTGATGTGTC 1259

273 GATGAGCGCGGAGAGC 288  
1260 GAAGTCTCTGAGAGC 1275

RESULT 13  
US-10-716-803-4  
Sequence 4, Application US/10716803  
Publication No. US20040229236A1  
GENERAL INFORMATION:  
APPLICANT: Blanc, Veronique  
Blanche, Francis  
Crouzet, Joel  
Jacques, Nathalie  
Lacroix, Patricia  
Thibaut, Denis  
Zagorec, Monique  
Debussche, Laurent  
De Crecy-Lagard, Valerie  
TITLE OF INVENTION: Polypeptides Involved In The  
Biosynthesis Of Streptogramins, Nucleotide Sequences  
Coding For These Polypeptides And Their Use  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
City: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/716,803  
FILING DATE: 20-Nov-2003  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/635,359  
FILING DATE: 09-AUG-2000  
APPLICATION NUMBER: US 09/231,818  
FILING DATE: 15-JAN-1999  
APPLICATION NUMBER: US 08/403,852  
FILING DATE: 10-MAY-1995  
APPLICATION NUMBER: PCT/FR 93/00923  
FILING DATE: 25-SEP-1993

APPLICATION NUMBER: FR 92/11441  
FILING DATE: 25-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Arrigo, Salvatore J.  
REGISTRATION NUMBER: 46,063  
REFERENCE/DOCKET NUMBER: 03806.0054-04000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1208 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: S.priestinaespiralis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1208  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-10-716-803-4

Query Match 17.3%; Score 54; DB 20; Length 1208;  
Best Local Similarity 50.0%; Pred. No. 3.4e-05;  
Matches 135; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

38 AGCTGACGCGCTCGACGGGAGCCTCTTGACACGCTGCGGCGCGCATCGACTCGGTG 97  
860 AGGTGACCGCTTCGCGCGCGGTACGCGATGCTGGGTCCCAAGAACTGTCGCGCGG 919

98 TCCGATCGCGGGGTACAGTCCCGGACAGGGGTCCCATGATGACAGCCCGCGGTGA 157  
920 GCTCGCGCTCCGCTCGAGGTTCAGGTGCTTACGCCATTCGCAAGGCGGAGCGGTG 979

158 GCCTGTCAAGAGCAGGCGCGCGCTACGCGCGGACCGGCTTCGACGATCGTTC 217  
980 GCTGTTCGTGACAGCTTCGCGACCGGACCGGCTGCGCCAGAGCGCATCGAAGGCCA 1039

218 TGGTGAACCTTACGACGATCATCAGAGATGTGCGCGCTCGAGGACCTGTGATGA 277  
1040 TCACGAGGTCCTTCGACCTGCGCGCGCGCGCATATCGCGACCTCGACTGTGCGGC 1099

278 GCGGAGAGCTGACGCGCGCGGAGACCGG 307  
1100 CCATCTACGCGCGCACCGCGCGCTTACGGCC 1129

RESULT 14  
US-10-716-803-1  
Sequence 1, Application US/10716803  
Publication No. US20040229236A1  
GENERAL INFORMATION:  
APPLICANT: Blanc, Veronique  
Blanche, Francis  
Crouzet, Joel  
Jacques, Nathalie  
Lacroix, Patricia  
Thibaut, Denis  
Zagorec, Monique  
Debussche, Laurent  
De Crecy-Lagard, Valerie  
TITLE OF INVENTION: Polypeptides Involved In The  
Biosynthesis Of Streptogramins, Nucleotide Sequences  
Coding For These Polypeptides And Their Use  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
City: Washington



```
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/716,803
FILING DATE: 20-Nov-2003
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 09/635,359
FILING DATE: 09-AUG-2000
APPLICATION NUMBER: US 09/231,818
FILING DATE: 15-JAN-1999
APPLICATION NUMBER: US 08/403,852
FILING DATE: 10-MAY-1995
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Arriago, Salvatore J.
REGISTRATION NUMBER: 46,063
REFERENCE/DOCKET NUMBER: 03806.0054-04000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5392 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-716-803-1

Query Match      17.3%; Score 54; DB 20; Length 5392;
Best Local Similarity 50.0%; Pred. No. 3.2e-05;
Matches 135; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 38 AGCTCGACGCGCTCGACGCGCTCTCTGACACGCGTGGCGCGCGCGCATCGACCTCGGTG 97
DB 4417 AGGTCGACCGCTTCGCGCGCGCTGACGCGATGCGCTGGGTCGCCAAGAACGTCGCGCGG 4476
QY 98 TCCGATCGCGCGGTACAGATCCCGGACGCGCGTCCGATGATGACGCCGCGCGGTCA 157
DB 4477 GCGTCGCGCTCCGCGTGGAGGTCCAGGTCCGCTACGCGCATCGGACGCGAGCGGTG 4536
QY 158 GCGTGGTCAAGACGAGGCGCGCGCTACGCGCGCGACGCGCGCTGACGAATGTTCC 217
DB 4537 GCGTGTTCGTGAGACCTTCGCGCACCGCGCACCGTCCGCCAGAGCGCATGAGAGGCCA 4596
QY 218 TGGTGAACCTCTGACGATGATCATGACGAGATGTGCGCGTGAAGACCTGCTGATGA 277
DB 4597 TCACCGAGGTCTTGCACCTGCGCGCGCGCGCATCATCGGACCTGACCTGCTGCGGC 4656
QY 278 GCCGGGAGAGCTGACGCGCGCGAGACCGGC 307
DB 4657 CCATCTACGCGCGACCGCGCGCTACGCGCC 4686
```

```
RESULT 15
US-10-107-431-112
; Sequence 112, Application US/10107431
; Publication No. US20030224364A1
; GENERAL INFORMATION:
```

```
APPLICANT: Farnet, Chris
APPLICANT: Staffa, Alfredo
APPLICANT: Zazopoulos, Emmanuel
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTH
FILE OF INVENTION: BIOSYNTHETIC LOCI
FILE REFERENCE: 3001-7US
CURRENT APPLICATION NUMBER: US/10/107,431
CURRENT FILING DATE: 2002-03-28
NUMBER OF SEQ ID NOS: 282
SOFTWARE: Patent version 3.0
SEQ ID NO 112
LENGTH: 837
TYPE: DNA
ORGANISM: Streptomyces mobaraensis
US-10-107-431-112

Query Match      17.2%; Score 53.6; DB 17; Length 837;
Best Local Similarity 52.1%; Pred. No. 4.5e-05;
Matches 147; Conservative 0; Mismatches 129; Indels 6; Gaps 1;

QY 28 CTGCGGCGGAGCTCGACGCGCTCTGACGCGGACGCTCTTGACACGCTGCGCGCGCATC 87
DB 151 CTGCTGACCTCCGCGACGCGCTGCGCGGAGCTCCGCGCGCGCTCCGCGCGCTGCTC 210
QY 88 GACCTGCGGTCTCGCATCGCGCGGTACAGTCCCGGACGCGCGTCCCGATGATGACGCC 147
DB 211 GACTTGGCGCGCGCGACGTCGCTACCGCGCTGCTCTCCGCGCGCGATGAGACC 270
QY 148 GCGCGGATGACCTGCTCAAGACGAGGCGCGCGCTACGCGCGCGACGCGCGCTGAC 207
DB 271 GCGGATGCGCGCGCGCGCGGACGACCTGACCGCGCGACGACGAGTGAACGCGCGCTG 330
QY 208 GAATGCTTCTGTGTAACCTCTACGACGTGATCATGACGAGATGTGCGCGCTGAGAC 267
DB 331 AGCGCCTGCGGACGCGCTGCTGACGCGGATCTCTC-----CACCGAGTCTGAG 384
QY 268 CTGCTGATGACCGCGGAGAGCTGACGCGCGCGAGACCGCGCG 309
DB 385 CACGTACGAGACCGGACACCGACCTGCGGAGGCGCTCCGG 426
```

Search completed: October 5, 2005, 09:57:04.  
Job time : 331.305 secs



This Page Blank (uspto)







Db 910 GAATTAGGCGCAGCTCAGCCGACTTCATCATCCCGGACCTGTGGAGAGTACGCTCAC 969  
QY 136 ATGATGACAGCCCGCGGCTGACCTGTGTCAGACAGAGGCGCGCTTACCGCGCGAC 195  
Db 970 GTCAGAGCTACGTCAGAGTCAACCTTCATCAGAGGCCAGAGATCTCTCTCTTCGAC 1029  
QY 196 CACGCGCTCGACGAATGTTCTGTGAACCTCTGACGCTGATCATCAGAGATGTGC 255  
Db 1030 GTGCGGCTGGCGCAGTACGCCACACCTCTCCAAAGTACGCGCTGAACCTGTGCGC 1089  
QY 256 CCGGTGAGAGACCTGTGTATGAGCCGGAGAGCCTTACGCGCGAGAGCGCGC 308  
Db 1090 GCGGTGTGAAGAGGTGAAGCGGAGAGATGAGACTGACGACGCGAGCGC 1142

RESULT 2  
CNA34658 422 bp mRNA linear EST 07-JUL-2004  
LOCUS BE030008B10H11 BE03 Normalized and Subtracted bovine embryonic and  
DEFINITION extreme embryonic tissue Bos taurus cDNA clone BE030008B10H11 5', mRNA  
sequence.  
CNA34658  
VERSION CN434658.1 GI:46413922  
KEYWORDS EST.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovinae; Bos.

REFERENCE 1 (bases 1 to 422)  
AUTHORS Lewin,H.A., Renard,J.P., Yang,X.J., Hernandez,A., Degrelle,S.,  
Hue,I., Tian,X.C., Liu,L. and Everts,R.E.  
TITLE Bovine embryonic ESTs  
JOURNAL Unpublished (2004)  
COMMENT Contact: Harris Lewin  
Department of Animal Sciences  
University of Illinois at Urbana-Champaign  
206 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL  
61801, USA  
Tel: 217 333 5998  
Fax: 217 244 5617  
Email: h-lewin@uiuc.edu

Funding for embryonic EST sequencing was provided by USDA ARS  
contract No. 58-1275-2-020 to H. A. Lewin. Base Calling/Quality  
scores: PHRED from Washington University Genome Center. Vector  
Trimming: Cross\_match from Washington University Genome Center.  
PHRAP site. Sequences submitted are vector free and at least 200  
bp in length with average PHRED score > 20. REPEAT IN THE SEQUENCE  
low\_complexity STRAND (+) ELEMENT GC\_rich LOCATION [350,363].  
PCR primers  
FORWARD: CCGATCAGACGTTGTAAACGAC (M13 NIA-F)  
BACKWARD: GTGTGATTTGAGCGATACAC (M13 NIA-R)  
Insert Length: 422 Std Error: 0.00  
Plate: BE030008B10 row: H column: 11  
Seq primer: TAATACGACTCACTATAGG (T7 PROMOTER)  
High quality sequence stop: 422.  
Location/Qualifiers

FEATURES  
source 1..422  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/clone="BE030008B10H11"  
/issue\_type="embryo (day 36 and day 64) and  
extra-embryonic tissue (day 14 to 25) Normalized and  
subtracted library"  
/lab\_host="DH10B"  
/clone\_lib="BE03 Normalized and Subtracted bovine  
embryonic and extraembryonic tissue"  
/note="Vector: pGEMZf1(+); Site 1: EcoRI; Site 2: NotI;  
The cDNA library was constructed, normalized and  
subtracted by Dr. A. Hernandez, W.M. Keck Center,  
University of Illinois Urbana-Champaign and S. Degrelle,

BDR, Institut National de la Recherche Agronomique,  
France, as described in Bonaldo, M.F., Lemon, G. and  
Soares, M.B. (1996), Genome Research 6(9): 791-806. This  
clone is from library BE03, a normalized library of  
embryonic tissues substracted with 7800 known placental  
sequences. The double stranded cDNA was size selected  
(more than 450 bp), adapted with EcoRI adaptors at both  
ends and then digested with NotI. The cDNA was then  
directionally cloned into EcoRI-NotI digested pGEM11zf(+)  
plasmid vector. Insert size was between 450-2200 bp.  
Sequencing was done from the 5' end of the clone. Tagged  
oligo-dt was used to identify the source of the ESTs where  
possible. (A18)TCGT = extreme embryonic tissue; (A18)TCGT =  
36-day old fetus; (A19)TCGA = 64-day old fetus (organs);  
(A18)TCGA = 64-day old fetus (body)"

ORIGIN  
Query Match 17.5%; Score 54.6; DB 7; Length 422;  
Best Local Similarity 51.0%; Pred. No. 0.071;  
Matches 129; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 1 ATGACCGACGACGAGAGCTGACGCGGCTGCGGAGCTGACGCGCTCGACGCGAGC 60  
Db 18 ATGATGCGGACGACCGTGAACGCGCTGTGCGGCGGTGCTTACCTTCACACCGTG 77  
QY 61 CTCTTGACACGCGTGGCGCGCCGATCGACTCGGTGTCGCGATCGCGGCTTACAGTCC 120  
Db 78 CCGCGCGCATATATGATGATGATGCGGCGCGCAAGGAGCGGACATATATGACAAAGGCGAC 137  
QY 121 CCGACGCGGCTGCCGATGATGACGCGCGCGGCTGACGCGGCTTCTTCAAGGCGCGC 180  
Db 138 GTGACGTGTGCGGAGAGTCTTCAAGGAGCGGCGCGGAGGCGCTTCTTCAAGGCGCGC 197  
QY 181 CGCTACGCGCGCGACGACGCGCGCTCGACGAATGTTCTGTGTAACCTTACGACGATGC 240  
Db 198 TGTCTCAACGTTCTGCGCGGCGCATGCGGCGCGCTTCTGTCTCTTCAACGACGCTC 257  
QY 241 ATCAGCGAGATGT 253  
Db 258 AAGAGGTATCT 270

RESULT 3  
AV610217/c 653 bp mRNA linear EST 28-NOV-2001  
LOCUS AV610217 Bos taurus lung fetus Bos taurus cDNA clone E1U031A01 3',  
DEFINITION mRNA sequence.  
ACCESSION AV610217 GI:9745887  
VERSION AV610217  
KEYWORDS EST.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovinae; Bos.  
REFERENCE 1 (bases 1 to 653)  
AUTHORS Takasugi,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.  
TITLE Establishment of a high throughput EST sequencing system using  
poly(A) tail-removed cDNA libraries and determination of 36,000  
bovine ESTs  
JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)  
MEDLINE 21570554  
PubMed 11713328  
COMMENT Contact: Yoshikazu Sugimoto  
Animal Genetics Division  
Shirakawa Institute of Animal Genetics  
Odakura, Nishigo, Nishio-shirakawa, Fukushima 961-8061, Japan  
Tel: 81-248-25-5641  
Fax: 81-248-25-5725  
Email: kazusugi@cocoa.com.ne.jp  
Single pass sequencing.  
This clone was obtained from a polyA-deleted cDNA library.







```

High quality sequence stop: 678
FEATURES
Location/Qualifiers
source
1..678

```



The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones that display the same hybridisation matrix with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per ONFP cluster was selected for sequencing. cDNA clones and filters are distributed via Deutsches Ressourcenzentrum fuer Genomforschung GmbH (<http://www.rzpd.de>).

PCR Primers  
 FORWARD: 5' CCCAGGCTTACCACTTATCCGGCTCG 3' (M13SP) 5'-seq  
 BACKWARD: 5' GCTATTACGCGACGTGGCAAGGGGATGTC 3' (M13SP) 3'-seq  
 Seq primer: 5' GCTATTACGCGACGTGGCAAGGGGATGTC 3' (M13SP).

## FEATURES

source

```

1..712
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="R2PDI056U1856Q"
/sex="female"
/tissue_type="brain tissue"
/dev_stage="adult brain"
/clone_1lb="normal cattle brain"
/note="Organ: Brain; Vector: pSport1; Site_1: NotI; Site_2: SalI; Random primed and directionally cloned in pSport1 vector using NotI (5'-PGACTAGTCTAGTACGCGAGCGCGGCC (T)15-3' and SalI 5'-TCGACCCAGCGCTCCG-3' adapters (Gibco BRL)."

```

## ORIGIN

Query Match 17.5%; Score 54.6; DB 7; Length 712;  
 Best Local Similarity 51.0%; Pred. No. 0.069;  
 Matches 129; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

```

QY 1 ATACCGAGGAGAACGAGCTGACGCGCTCGCGCGAGCTCAACCCCTCGACGGAGC 60
DB 659 ATATATGCGCGAGACCGTGAAGCGCGCGGTGCTCTTCAACCTTCGACACCGTG 600
QY 61 CTCCTGAGACGAGTGGCGCGCGCATCGACCTCGGTGTCATCGCGGAGTTC 120
DB 599 CGCGCGCGCATATGATGACGTGGCGCGCGCAAGAGCCGACATCATGACAGGCAACC 540
QY 121 CGGACGCGCGTCCCGATGATGACGCCGCGCGGTGACCTGCTCAAGACAGGCGCGCC 180
DB 539 GTGAGCTGCTGGCGGAGATCTCAAGAGAGAGGCGGCAAGCCCTTCGAGGCGCC 480
QY 181 CGCTAGCGCGCGACCAACGCGCTTCGACGAATGCTTCTGTGAACCTTACGAGTGATC 240
DB 479 TGTCTCAAGCTTGTGGCGCGCATGGGCGCGCTGCTGCTGCTCTTCAAGACGAGCTC 420
QY 241 ATCAGGAGATGT 253
DB 419 AAGAGGTCTCT 407

```

## RESULT 8

CN436828

LOCUS

DEFINITION

714 bp mRNA linear EST 07-JUL-2004  
 BE04010B1F12 BE04 Normalized and Subtracted bovine embryonic and extraembryonic tissue Bos taurus cDNA clone BE04010B1F12 5', mRNA sequence.

ACCESSION CN436828  
 VERSION CN436828  
 KEYWORDS EST.  
 SOURCE CN436828.1 GI:46416092

ORGANISM Bos taurus (cow)  
 Bos taurus

Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 714)

AUTHORS Lewin,H.A., Renard,J.P., Yang,X.J., Hernandez,A., Degrelle,S., Hue,I., Tian,X.C., Liu,L. and Everts,R.E.

TITLE Bovine embryonic ESTs  
 JOURNAL Unpublished (2004)  
 COMMENT Contact: Harris Lewin

Department of Animal Sciences  
 University of Illinois at Urbana-Champaign  
 206 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL 61801, USA  
 Tel: 217 333 5998  
 Fax: 217 244 5617  
 Email: h-lewin@uiuc.edu

Funding for embryonic EST sequencing was provided by USDA ARS contract No. 58-1275-2-020 to H. A. Lewin. Base Calling/Quality Scores: PHRED from Washington University Genome Center. Vector Trimming: Cross match from Washington University Genome Center PHRAP suite. Sequences submitted are vector free and at least 200 bp in length with average PHRED score > 20. REPEAT IN THE SEQUENCE low complexity STRAND (+) ELEMENT GC-rich LOCATION [385,418].

PCR PRIMERS  
 FORWARD: CCAATCAGCAGCTTGAAGAAGC (M13 NTA-F)  
 BACKWARD: GTGTGGAATGTGACCGGATACAA (M13 NTA-R)

Insert Length: 714 Std Error: 0.00  
 Plate: BE04010B1 row: F column: 12  
 Seq primer: TATAGACTCACTATAGG (T7 PROMOTER)  
 High quality sequence stop: 714.

## FEATURES

source

```

1..714
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="BE04010B1F12"
/tissue_type="embryo (day 36 and day 64) and extra-embryonic tissue (day 14 to 25) Normalized and twice subtracted library"
/lab_host="DH10B"

```

/clone\_1lb="BE04 Normalized and Subtracted bovine embryonic and extraembryonic tissue"  
 /note="Vector: pGEMZ11(+); Site 1: EcoRI; Site 2: NotI; The cDNA library was constructed, normalized and subcloned by Dr. A. Hernandez, W.M. Reck Center, University of Illinois Urbana-Champaign and S. Degrelle, BPR, Institut National de la Recherche Agronomique, France, as described in Bonaldi, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. This clone is from library BE04, a normalized library of embryonic tissues subcloned with 7800 known placental sequences and 5000 clones from the BE03 library. The double stranded cDNA was size selected (more than 450 bp), adapted with EcoRI adaptors at both ends and then digested with NotI. The cDNA was then directionally cloned into EcoRI-NotI digested pGEMZ11F(+) phagemid vector. Insert size was between 1450-2200 bp. Sequencing was done from the 5' end of the clone. Tagged oligo-dT was used to identify the source of the ESTs where possible. (A18)TGGCT = extraembryonic tissue/ (A18)TGGGT = 36-day old fetus; (A18)TGGCA = 64-day old fetus (organs); (A18)TCCGA = 64-day old fetus (body)"

## ORIGIN

Query Match 17.5%; Score 54.6; DB 7; Length 714;  
 Best Local Similarity 51.0%; Pred. No. 0.069;  
 Matches 129; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

```

QY 1 ATACCGAGGAGAACGAGCTGACGCGCTCGCGCGAGCTCAACCCCTCGACGGAGC 60
DB 47 ATATATGCGCGAGACCGTGAAGCGCGGTGCTCTTCAACCTTCGACACCGTG 106
QY 61 CTCCTGAGACGAGTGGCGCGCGCATCGACCTCGGTGTCATCGCGGAGTTC 120
DB 107 CGCGCGCGCATATGATGACGTGGCGCGCGCAAGAGCCGACATCATGACAGGCAACC 166
QY 121 CGGACGCGCGTCCCGATGATGACGCCGCGGTGACCTGCTCAAGACAGGCGCGCC 180
DB 167 GTGAGCTGCTGGCGGAGATCTCAAGAGAGAGGCGCGCAAGCCCTTCGAGGCGCC 226
QY 181 CGCTAGCGCGCGACCAACGCGCTTCGACGAATGCTTCTGTGAACCTTACGAGTGATC 240

```



```

Db      227 TGCTCCACGTTTGGCGCGCATGGCGCGGCGCTTGCTGTGCTCTTACGACGAGCTC 286
QY      241 ATCAGCGAGATGT 253
        ||| ||| |||
Db      287 AAGAGGTCATCT 299

RESULT 9
CO882659      738 bp      mRNA      linear      EST 01-SEP-2004
LOCUS      BovGen.10984 normal cattle brain Bos taurus cDNA clone
DEFINITION      RZPD01056D1724Q 5', mRNA sequence.
ACCESSION      CO882659
VERSION      CO882659.1 GI:51812591
KEYWORDS      EST.
SOURCE      Bos taurus (cow)
ORGANISM      Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE      1 (bases 1 to 738)
AUTHORS      Hennig S., Janitz M., Herwig R. and Williams J.
TITLE      Generation, annotation, evolutionary analysis and database
            integration of 14869 cattle EST clusters
JOURNAL      Unpublished (2004)
COMMENT      Contact: Hennig S
            Laboratory 123, dept.Lehrach
            Max-Planck-Institut fuer Molekulare Genetik
            Ihnestr.63-73, D-14195 Berlin, Germany
            Tel: +49 30 8413 1612
            Fax: +49 30 8413 1380
            Email: hennig@molgen.mpg.de
            The library was characterised by oligonucleotide fingerprinting
            (ONFP) to reduce sequencing redundancy. According to the ONFP
            procedure, clones that display the same hybridisation matrix with a
            battery of 200 8mer oligonucleotides are grouped into clusters. One
            clone per ONFP cluster was selected for sequencing. cDNA clones and
            filters are distributed via Deutsches Ressourcenzentrum fuer
            Genomforschung GmbH (http://www.rzpd.de).
PCR PRIMERS
FORWARD: 5' CCCGAGCTTTACACTTATGCTTCGGGCTCG 3' (M13RSP) 5'-seq
BACKWARD: 5' GGTATTACGCGCGATCGCGAAGGGGAGATGTG 3' (M13RSP) 3'-seq
Seq primer: 5'-CCGGTCGGAATTCGCGGT-3' (M13RSP).
Location/Qualifiers
1. 738
   /organism="Bos taurus"
   /mol_type="mRNA"
   /db_xref="taxon:9913"
   /clone="RZPD01056D1724Q"
   /sex="female"
   /tissue_type="brain tissue"
   /dev_stage="adult brain"
   /clone_lib="normal cattle brain"
   /note="Organ: brain; Vector: pSport1; Site 1: NotI;
   Site 2: SalI; Random primed and directionally cloned in
   pSport1 vector using NotI
   (5'-pGACTGATCTAGATCGGATCGGAGCGCGCGCC (T)15-3' and SalI 5'-
   TCGACCCAGCGCTCG-3' adapters (Gibco BRL)."
ORIGIN
Query Match      17.5%; Score 54.6; DB 7; Length 738;
Best Local Similarity 51.0%; Pred. No. 0.069;
Matches 129; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY      1 ATGACGCGACGAGAGAGAGCTGACGCGCTGCGCGGAGCTGACGCCCTCGACGGGAGC 60
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      58 ATGATCGCGAGCGCTGACGCGCGCTGCGGCGGCTGCTCTTACCCCTTCGACACCGTG 117
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      61 CTCCTGAGACGCTGCGCGCGCGCATCGACCTTCGTGTCGCGATCGCGCGGTACAGTCC 120
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      118 CGGCGCGCGATATGATGACAGTGGGGCGCAAGAGCGGACATCATGTACAAGGGCACC 177
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      121 CGGACGCGGTCCGATGATGACGCCGCGCGGCTGCTGTAAGGACAGGGCGGCC 180
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db      178 GTGACTGTGCGGAGATCTCTCAAGACGAGGCGGCGAGGCCCTTCTTCAAGGGCGCC 237
QY      181 CGCTAGCGCCGACACGAGCGCTTCGACGAGATGTTCTGTGACCTTACGAGCGATGC 240
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      238 TGCTCCACGTTTCTGCGCGCATGGCGCGGCGCTTGCTGTGCTCTTACGACGAGCTC 297
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      241 ATCAGCGAGATGT 253
        ||| ||| |||
Db      298 AAGAGGTCATCT 310

RESULT 10
CR552094      740 bp      mRNA      linear      EST 12-JUL-2004
LOCUS      CR552094 Normalized and Subtracted bovine embryonic and
DEFINITION      extraembryonic tissues (bcaz) Bos taurus cDNA clone bcaz0007a.j.09
ACCESSION      CR552094
VERSION      CR552094.1 GI:50247004
KEYWORDS      EST.
SOURCE      Bos taurus (cow)
ORGANISM      Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE      1 (bases 1 to 740)
AUTHORS      Renard J.P., Lewin H.A., Yang J., Hernandez A., Degrelle S.,
            Everts R.E. and Hue I.
TITLE      Embryonic ESTs (bcaz)
JOURNAL      Unpublished (2004)
COMMENT      Contact: Renard JP
            Biologie du Developpement et Reproduction
            INRA
            Domaine de Vilvert 78350 Jouy en Josas, FRANCE
            Email: renard@jouy.inra.fr
            Funding for EST sequencing was provided by funds dedicated to
            J.P. RENARD (BDR, Inra). Funding for construction and subraction of
            the cDNA library was provided by USDA ARS contract No 58-1275-2-020
            to H. A. Lewin The cDNA library was constructed, normalized and
            substracted by Dr. A. Hernandez (W.M. Keck Center, University of
            Illinois Urbana-Champaign) and S. Degrelle (BDR, Inra) as described
            in Bonaldi et al. (1996), Genome Research 6, 791-806. BE03 library
            is a normalized library of embryonic tissues substracted with 7800
            known EST sequences. The double-stranded cDNA was size selected
            (>450 bp), adaptored with EcoRI adaptors at both ends and digested
            with NotI. The cDNA was then directionally cloned into EcoRI-NotI
            digested pGEMTIZ(+) phagemid vector. Insert size was between
            450-2200 bp. Sequence cleaned of vector, adaptor and repetitions.
            Contact us at sigenasupport@jouy.inra.fr to obtain the chromatogram
            of this sequence.
Location/Qualifiers
1. 740
   /organism="Bos taurus"
   /mol_type="mRNA"
   /db_xref="taxon:9913"
   /clone="bcaz0007a.j.09"
   /tissue_type="embryo (day 36 and day 64), extra-embryonic
   tissue (day 14 to 25)"
   /lab_host="DH10B"
   /clone_lib="Normalized and Subtracted bovine embryonic and
   extraembryonic tissues (bcaz)"
   /note="Vector: pGEMTIZ(1); Tagged oligo-dT was used to
   identify the origin of the ESTs where possible. (A18) TCGGT
   = extraembryonic tissue; (A18) TCGGT = 36-day old fetus;
   (A18) TCGGA = 64-day old fetus (organs); (A18) TCGGA =
   64-day old fetus (body)"
ORIGIN
Query Match      17.5%; Score 54.6; DB 7; Length 740;
Best Local Similarity 51.0%; Pred. No. 0.063;
Matches 129; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

```



QY 1 ATGACCGAGCAGAACGAGCTGACCGGCTGCGCGGAGCTGACGCGCTTGACGGGACG 60  
Db 300 ATGATCGCGGAGACCGTAGAGCGCGGTGGCGGGCGGTGCTCTTACCCCTTGACACCGTG 359  
QY 61 CTCTTGACACGCGTGGCGCGCGCATGCACTCGGTGTCCGCATCGCGGTACAGTCC 120  
Db 360 CGGCGGGCGCATGATGATGCACTCGGGGCGCAAGAGCGCATCATCTACAGAGGCGACC 419  
QY 121 CGGACGCGCGTCCGATGATGACAGCCGCGCGGCTGACGCTGTCAGAGACAGGGCGGCC 180  
Db 420 GTGACTGCTGGCGGAAAGATCTTCAAGAGACGAGGCGCGCAAGGCTTCTTCAAGGGCGCC 479  
QY 181 CGTACGCGCGCGCAGCAGCGGCTCGACGAATCGTTCGTGTGAACTCTACAGAGTATC 240  
Db 480 TGGTCCAAAGCTTTCGCGCGGATGGCGGCGGCTTCTGCTGTGCTCTTACAGAGAGCTC 539  
QY 241 ATCAGCGAGATGT 253  
Db 540 AAGAGGTCATCT 552

RESULT 11  
COB874118 745 bp mRNA linear EST 01-SEP-2004  
LOCUS BOvGen 02443 normal cattle brain Bos taurus cDNA clone  
DEFINITION R2PDP1056J1856Q 5', mRNA sequence.  
ACCESSION COB874118  
VERSION COB874118.1 GI:51803967  
KEYWORDS EST.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus

REFERENCE Hemmig, S., Janitz, M., Herwig, R. and Williams, J.  
AUTHORS Generation, annotation, evolutionary analysis and database  
TITLE Integration of 14969 cattle EST clusters  
JOURNAL Unpublished (2004)  
COMMENT Contact: Hemmig S  
Laboratory 123, dept. Lehnach  
Max-Planck-Institut fuer Molekulare Genetik  
Inmestr.63-73, D-14195 Berlin, Germany  
Tel: +49 30 8413 1612  
Fax: +49 30 8413 1380  
Email: hemmig@molgen.mpg.de

The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones that display the same hybridisation matrix with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per ONFP cluster was selected for sequencing. cDNA clones and libraries are distributed via Deutsches Ressourcenzentrum fuer Genomforschung GmbH (<http://www.rzpd.de>).  
PCR Primers  
FORWARD: 5' CCCAGGCTTACACTTATGCTCGGCGTGG 3' (M13RSP) 5'-seq  
BACKWARD: 5' GCTATTACCGCAGCTGCGCAAGAGGAGTGTG 3' (M13FSP) 3'-seq  
Seq primer: 5'-CCGATCGGAGTTCGCGGT-3' (M13RSP).  
Location/Qualifiers

FEATURES  
source  
1. 745  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/clone="R2PDP1056J1856Q"  
/sex="Female"  
/tissue\_type="brain tissue"  
/dev\_stage="adult brain"  
/clone\_lib="normal cattle brain"  
/note="Organ: brain; Vector: pSport1; Site 1: NotI;  
Site 2: SalI; Random primed and directionally cloned in  
pSport1 vector using NotI  
(5'-pGACTAGTCTAGATCGCGGCGCGCGCC (T)15-3' and SalI 5'-  
TCGACCCACCGCTCGG-3' adapters (Gibco BRL))"

ORIGIN

Query Match 17.5%; Score 54.6; DB 7; Length 745;  
Best Local Similarity 51.0%; Pred. No. 0.069;  
Matches 129; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 1 ATGACCGAGCAGAACGAGCTGACCGGCTGCGCGGAGCTGACGCGCTTGACGGGACG 60  
Db 466 ATGATCGCGGAGACCGTAGAGCGCGGTGGCGGGCGGTGCTCTTACCCCTTGACACCGTG 525  
QY 61 CTCTTGACACGCGTGGCGCGCGCATGCACTCGGTGTCCGCATCGCGGTACAGTCC 120  
Db 526 CGGCGGGCGCATGATGATGCACTCGGGGCGCAAGAGCGCATCATCTACAGAGGCGACC 585  
QY 121 CGGACGCGCGTCCGATGATGACAGCCGCGCGGCTGACGCTGTCAGAGACAGGGCGGCC 180  
Db 586 GTGACTGCTGGCGGAAAGATCTTCAAGAGACGAGGCGCGCAAGGCTTCTTCAAGGGCGCC 645  
QY 181 CGTACGCGCGCGCAGCAGCGGCTCGACGAATCGTTCGTGTGAACTCTACAGAGTATC 240  
Db 646 TGGTCCAAAGCTTTCGCGCGGATGGCGGCGGCTTCTGCTGTGCTCTTACAGAGAGCTC 705  
QY 241 ATCAGCGAGATGT 253  
Db 706 AAGAGGTCATCT 718

RESULT 12  
CR553143 754 bp mRNA linear EST 12-JUL-2004  
LOCUS CR553143 Normalized and Subtracted bovine embryonic and  
DEFINITION extraembryonic tissues (bcaz) Bos taurus cDNA clone bcaz0010a.m.23  
ACCESSION CR553143  
VERSION CR553143.1 GI:50248053  
KEYWORDS EST.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus

REFERENCE Renard, J.P., Lewin, H.A., Yang, J., Hernandez, A., Degrelle, S.,  
AUTHORS Everts, R.E. and Hue, I.  
TITLE Embryonic ESTs (bcaz)  
JOURNAL Unpublished (2004)  
COMMENT Contact: Renard JP  
Biologie du Developpement et Reproduction  
INRA  
Domaine de Vilvert 78150 Jouy en Josas, FRANCE  
Email: renard@jouy.inra.fr  
Funding for EST sequencing was provided by funds dedicated to  
J.P. RENARD (BDR, Inra). Funding for construction and subtraction of  
the cDNA library was provided by USDA ARS contract No 58-1275-2-020  
to H. A. Lewin. The cDNA library was constructed, normalized and  
subtracted by Dr. A. Hernandez (W.M. Keck Center, University of  
Illinois Urbana-Champaign) and S. Degrelle (BDR, Inra) as described  
in Bonafide et al. (1996), Genome Research 6, 791-806. B503 library  
is a normalized library of embryonic tissues subtracted with 7800  
known EST sequences. The double-stranded cDNA was size selected  
(>450 bp), adapted with EcoRI adaptors at both ends and digested  
with NotI. The cDNA was then directionally cloned into EcoRI-NotI  
digested pGEM1Zf(+) phagemid vector. Insert size was between  
450-2200 bp. Sequence cleaned of vector, adaptor and repetitions.  
Contact us at [sigenasupport@jouy.inra.fr](mailto:sigenasupport@jouy.inra.fr) to obtain the chromatogram  
of this sequence.  
Location/Qualifiers

FEATURES  
source  
1. 754  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/clone="bcaz0010a.m.23"  
/tissue\_type="embryo (day 36 and day 64), extra-embryonic  
tissue (day 14 to 25)"











This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 05:31:11 ; Search time 1525.07 Seconds  
(without alignments)  
3272.568 Million cell updates/sec

Title: US-10-089-514-4  
Perfect score: 514  
Sequence: 1 MTEQNELQRLAEALDGT.....MCRVEDLVMSRESLTADRR 103

Scoring table: BLOSUM62  
Xgapop 10.0, Ygapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 470823 segs, 24227607955 residues  
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seg length: 0  
Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USFTO.spool\_n/US10089514/runat\_04102005\_105744\_8033/app.query.fasta\_1.1621  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blomsm62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pcc -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10089514.@CCN\_1.1.8782.@runat\_04102005\_105744\_8033 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB	ID Description
1	514	100.0	312 6 BD178314 Transform
2	514	100.0	312 6 BD093915 Transform
3	514	100.0	5251 1 AB116234 Streptomy
4	486.5	94.6	14159 1 AF262220 Streptomy

5	256	49.8	387	6	A48325	A48325 Sequence 3
6	256	49.8	387	6	AR198355	AR198355 Sequence
7	256	49.8	2888	6	A48323	A48323 Sequence 1
8	256	49.8	2888	6	AR198353	AR198353 Sequence
9	256	49.8	4740	1	SP060417	SP060417 Streptomyce
10	120	23.3	349944	1	BX571871	BX571871 Phototrab
11	120	23.3	349980	6	AX770909	AX770909 Sequence
12	114	22.2	300454	6	AE017175	AE017175 Porphyrom
13	112	21.8	10632	1	U67480	U67480 Methanocald
14	112	21.8	110000	6	AR271569_12	Continuation (13 o
15	109.5	21.3	13242	1	AE000858	AE000858 Methanoba
16	109	21.2	4927	1	AE010450	AE010450 Methanopy
17	107.5	20.9	110000	1	AE017180_19	Continuation (20 o
18	106	20.6	1135	6	AR376132	AR376132 Sequence
19	105	20.4	1305	6	AR387774	AR387774 Sequence
20	104	20.2	21387	1	AE008821	AE008821 Salimone1
21	104	20.2	274050	1	AL627276	AL627276 Salimone1
22	104	20.2	300029	1	AE016842	AE016842 Salimone1
23	103.5	20.1	301630	1	AE016942	AE016942 Bacteroid
24	102	19.8	1152	6	AX457135	AX457135 Sequence
25	102	19.8	3434	1	EHPEAGN	X60420 Erwinia herb
26	102	19.8	3434	1	ERMP7A	M74133 Erwinia her
27	101	19.6	1253	6	108481	108481 Sequence 3
28	101	19.6	2058	6	108480	108480 Sequence 1
29	101	19.6	3659	6	108487	108487 Sequence 15
30	101	19.6	4190	6	108488	108488 Sequence 19
31	101	19.6	4509	1	ECOPHEAB	M10431 E.coli phe
32	101	19.6	10327	1	AE015281	AE015281 Shigella
33	101	19.6	12071	1	AE005490	AE005490 Escherich
34	101	19.6	16772	1	D90887	D90887 E.coli geno
35	101	19.6	19338	1	D90888	D90888 E.coli geno
36	101	19.6	110000	1	U00096_27	Continuation (28 o
37	101	19.6	270355	1	AP002562	AP002562 Escherich
38	101	19.6	290380	1	AE016987	AE016987 Shigella
39	101	19.6	300039	1	AE016764	AE016764 Escherich
40	100.5	19.6	16905	1	AE005050	AE005050 Halobacte
41	100	19.5	11069	1	AE013693	AE013693 Yersinia
42	100	19.5	110000	1	BX936398_09	Continuation (10 o
43	100	19.5	110000	1	BX936398_10	Continuation (11 o
44	100	19.5	220050	1	AJ414156	AJ414156 Yersinia
45	100	19.5	290803	1	AE017128	AE017128 Yersinia

## ALIGNMENTS

RESULT 1	BD178314	312 bp	DNA	linear	PAT 16-APR-2003
LOCUS	BD178314	Transformant producing p1022 substance and process for producing the same and novel biosynthetic gene.			
DEFINITION	BD178314	the same and novel biosynthetic gene.			
ACCESSION	BD178314.1	GI:30015579			
VERSION	BD178314.1	GI:30015579			
KEYWORDS	WO 02077244-A/2.	Streptomyces venezuelae			
SOURCE	WO 02077244-A/2.	Streptomyces venezuelae			
ORGANISM	Streptomyces venezuelae	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomyces			
REFERENCE	1 (bases 1 to 312)	Streptomyces			
AUTHORS	Yanai,K., Sumida,N., Watanabe,M., Moriya,T. and Murakami,T.	Transformant producing p1022 substance and process for producing the same and novel biosynthetic gene			
TITLE	Patent: WO 02077244-A 2 03-OCT-2002;	MEIJI SEIKA KAISHA LTD,KOJI YANAI,NAOMI SUMIDA,MANABU WATANABE,TATSUKI MORIYA,TAKESHI MURAKAMI			
JOURNAL	OS Streptomyces venezuelae	PAT 16-APR-2003			
COMMENT	PN WO 02077244-A/2	OS Streptomyces venezuelae			
	PD 03-OCT-2002	PD 03-OCT-2002			
	PR 22-MAR-2002	PR 22-MAR-2002			
	PI KOJI YANAI,NAOMI SUMIDA,MANABU WATANABE,TATSUKI MORIYA,TAKESHI MURAKAMI	PI KOJI YANAI,NAOMI SUMIDA,MANABU WATANABE,TATSUKI MORIYA,TAKESHI MURAKAMI			
	PC C12N15/61,C12N1/15,C12P17/14/(C12N1/15,C12R1:645), (C12P17/14,	C12N15/61,C12N1/15,C12P17/14/(C12N1/15,C12R1:645), (C12P17/14,			



PC C12R1:645)  
CC Transformant producing PF1022 substance and process for CC  
CC and novel biosynthetic gene  
FH Key Location/Qualifiers  
FT CDS (1)..(309).  
FEATURES  
source 1..312  
/organism="Streptomyces venezuelae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:54571"  
ORIGIN  
Alignment Scores:  
Pred. No.: 9, 73e-47 Length: 312  
Score: 514.00 Matches: 103  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0  
US-10-089-514-4 (1-103) x BD178314 (1-312)  
QY 1 MetTnGluGlnAsnGluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20  
DB 1 ATGACCGAGACAGACGAGCTGCAGCGGCTGCCGCGGAGCTGCAGCCCTCGACGGGAGC 60  
QY 21 LeuLeuAspThrValAlaArgArgAlaLeuAspLeuGlyValArgAlaArgTyrLysSer 40  
DB 61 CTCCTGGACACGAGTGGGGCGCCGACATCGACCTCGGTGTCCGATCCGCGGTACAAAGTCC 120  
QY 41 ArgHisGlyValProMetMetGlnProGlyArgValSerLeuValLysAspArgAla 60  
DB 121 CGGACGGGGCTCCCATGATGACGCCGCCGCGGTACGCTGTCAAGACAGGCGCC 180  
QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIle 80  
DB 181 CGCTAGCGCGCGACACGAGCCTCGACGATCGTTCTCGTGTGAACCTTACGACGTGATC 240  
QY 81 IleThrGluMetCysArgValAlaGluAspLeuValMetSerArgGluSerLeuThrAlaGlu 100  
DB 241 ATCACGGAATGTGGCGCTCGAGGACCTGTGATGAGCGGAGAGCCTGACGGCCGAG 300  
QY 101 AspArgArg 103  
DB 301 GACCGGCGG 309  
RESULT 2  
BD093915 312 bp DNA linear PAT 27-AUG-2002  
LOCUS Transformants that produce secondary metabolites modified by a  
DEFINITION functional group(s) and novel biosynthesis genes.  
ACCESSION BD093915  
VERSION BD093915.1 GI:22639503  
KEYWORDS WO 0123542-A/2.  
SOURCE Streptomyces venezuelae  
ORGANISM Streptomyces venezuelae  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 312)  
AUTHORS Yanai,K., Okakura,K., Yaeuda,S., Watanabe,M., Miyamoto,K., Mido,N.  
and Murakami,T.  
TITLE Transformants that produce secondary metabolites modified by a  
functional group(s) and novel biosynthesis genes  
JOURNAL Meitai SEIKA KAISHA LTD,KOJI YANAI, KAORU OKAKURA,SHOHEI YASUDA,  
MANABU WATANABE,KOICHI MIYAMOTO,NAOKI MIDO, TAKESHI MURAKAMI  
OS Streptomyces venezuelae  
PN WO 0123542-A/2  
PD 05-APR-2001  
PF 29-SEP-2000 WO 2000JP006783  
PI 29-SEP-1999 JP 99P 276314  
KOJI YANAI,KAORU OKAKURA,SHOHEI YASUDA,MANABU WATANABE,KOICHI

PI MIYAMOTO,  
PI NAOKI MIDO,TAKESHI MURAKAMI  
PC C12N15/09,C12N5/10,C12P21/02,C07K11/00// (C12P21/02,C12R1:645)  
CC  
FH Key Location/Qualifiers  
FT CDS (1)..(309).  
FEATURES  
source 1..312  
/organism="Streptomyces venezuelae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:54571"  
ORIGIN  
Alignment Scores:  
Pred. No.: 9, 73e-47 Length: 312  
Score: 514.00 Matches: 103  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0  
US-10-089-514-4 (1-103) x BD093915 (1-312)  
QY 1 MetTnGluGlnAsnGluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20  
DB 1 ATGACCGAGACAGACGAGCTGCAGCGGCTGCCGCGGAGCTGCAGCCCTCGACGGGAGC 60  
QY 21 LeuLeuAspThrValAlaArgArgAlaLeuAspLeuGlyValArgAlaArgTyrLysSer 40  
DB 61 CTCCTGGACACGAGTGGGGCGCCGACATCGACCTCGGTGTCCGATCCGCGGTACAAAGTCC 120  
QY 41 ArgHisGlyValProMetMetGlnProGlyArgValSerLeuValLysAspArgAla 60  
DB 121 CGGACGGGGCTCCCATGATGACGCCGCCGCGGTACGCTGTCAAGACAGGCGCC 180  
QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIle 80  
DB 181 CGCTAGCGCGCGACACGAGCCTCGACGATCGTTCTCGTGTGAACCTTACGACGTGATC 240  
QY 81 IleThrGluMetCysArgValAlaGluAspLeuValMetSerArgGluSerLeuThrAlaGlu 100  
DB 241 ATCACGGAATGTGGCGCTCGAGGACCTGTGATGAGCGGAGAGCCTGACGGCCGAG 300  
QY 101 AspArgArg 103  
DB 301 GACCGGCGG 309  
RESULT 3  
AB116234 5251 bp DNA linear BCT 03-JUL-2004  
LOCUS Streptomyces venezuelae papa, papC, papB, ORFIV genes for  
DEFINITION 4-amino-4-deoxychorismate synthase, 4-amino-4-deoxyisoprenate  
dehydrogenase, 4-amino-4-deoxychorismate mutase, hypothetical  
protein, complete cds.  
ACCESSION AB116234  
VERSION AB116234  
KEYWORDS AB116234.1 GI:47846865  
SOURCE Streptomyces venezuelae  
ORGANISM Streptomyces venezuelae  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1  
AUTHORS Yanai,K., Sumida,N., Okakura,K., Moriya,T., Watanabe,M. and  
Murakami,T.  
TITLE Para-position derivatives of fungal antihelminthic cyclopeptides  
engineered with Streptomyces venezuelae antibiotic biosynthetic  
genes  
JOURNAL Nat. Biotechnol. 22 (7), 848-855 (2004)  
PUBMED 15184904  
REFERENCE 2 (bases 1 to 5251)  
AUTHORS Yanai,K., Sumida,N., Okakura,K., Moriya,T., Watanabe,M. and  
Murakami,T.  
TITLE Direct Submission







REMARK	SEQUENCE UPDATE BY SUBMITTER
COMMENT	On Jun 5, 2001 this sequence version replaced gi:10716943
FEATURES	Location/Qualifiers

**CDS**

```

/function="essential for chloramphenicol biosynthesis"
/codon_start=1
/transl_table=11
/product="CmlG"
/protein_id="AAG21974.1"
/db_xref="GI:10716944"
/translation="MTAGSGGAGGRTPLEPRALAVRDHSFLADGMDVDNPTAF
LVAEEDVVTLDPPGVEGGEALD
LVAAYHNSRPLRLIDEVRLAVNRHDDGEVLVFKSHLIVLDADVLDLRLRAHNG
DTPRGGIGSRYRHNAVTQYRLALPADPAETRLGELPGRKGIPITIASYEWML
PLEGTEYPELISAAGVSLGTRMADGGGLVQYYSWMEARVETLKIPLVPPGLG
SDTPEYFAKTRAAQSLSLORFTMSDLDVAFARFTSGEMRNVVSTTTEPEPRMR
VTPRSARVEDLKEFLADRSWMPGPILLRMQYKGGFSLPGAADVLDDGLKORIGASGGTPE

```

CDS

```

/note="non-ribosomal peptide synthetase"; contains
reductase-like domains near C-terminus";
/codon_start=1
/transeq_table=11
/product="CmHh"
/protein_id="FAG21975.2"
/db_xref="GI:143278997"
/transeq_location="MELPOLLIDLEPRRHGRYALRTAGRTWYEELODFVSLARRRIDLV
ACEQPARVRLVAGEHATVAGVWVWALAMSHAVHTPMNGLOPADREEEFAVDAADLLVY
CERFALVYGEKRGKALVAGVGMPTPPAPADGCTADEPARSRVANSITFSGSTGD
PKLVYDHGGLNLCKSLRRLLDTPDDVOYFALSLSTDAVSSEILGTVYAGATLVYVP
VRQASWLGSVSHLAAHCDCDLAMSPSYALALDPAASRLRKEVFCGEALGSEBYDR
AAVRSVFNAYGTEATVTFSELAETISYTPSISGTVDQFRAVYRPPSGDHATGEBE
LVYVGQGVLAAGTGVGSPANEEFVGVDSPPAATGVDSLDGELTVLGRIDBOIKRAI
LGRVNLVANGSTLSHGLKVRALYQODATILVYPAAGEAETEEELARVINDLVPMVE
APRVLVINALPLTSGCKDRSALRELLASPAAGHGCTDGEADAELRVLDVYTAIVL
GSAIPGSETSIFFAGSGSSIAMIQVCKLSAYAEBAEVAEAPAMQYDPRAPAFLLHILRG
EAPAPBSAVIDTLRVLVSTETBRDLRELPRLKRDTRHREAPYARADGEVLLTASGSE
IGHVLDRLLAARPELVVSTEDPGVLTGHAITRGQADAPYARVLAIVALELRWVD
RRGRGPAVAVVHCGQVNHLPIDSHLSSGVNITLVVRAAALGARSFALISAASGHL
ADLPALPSAATLPAVGQPSRSKLISEEYVNTLAVGCAVSHYRPLVYGHFDEPRHIL
KDMFRTALLETARVGMAMRSLGSHVCOVGLTADTLILGRPPANGTADGASRTPPDSR
SAVHYRATVLADELLRHNTLRTADVLVAPAAPEERIRDSGEVRLPILAMQALSSPCMK
PSAAREVHDHILGRLLGTPPTPPAGDRPEPRTFTISRGERRRRPRADTLTGTEP"
3576..3596
/note="Region: adenylation domain A1"
3708..3746
/note="Region: adenylation domain A2"
3939..3986
/note="Region: adenylation domain A3"
4089..4100
/note="Region: adenylation domain A4"
4350..4370
/note="Region: adenylation domain A5"
4494..4538
/note="Region: adenylation domain A6"
4593..4610
/note="Region: adenylation domain A7"
4647..4706
/note="Region: adenylation domain A8"
4833..4853
/note="Region: adenylation domain A9"
4893..4910
/note="Region: adenylation domain A10"
5043..5075
/note="Region: thiolation domain"
5376..5375
/note="Region: NADH binding site"

```

ORIGIN

### Alignment Scores:

Pred. No.:	5.2e-42	Length:	1155
Score:	486.50	Matches:	103
Percent Similarity:	98.10%	Conservative:	0
Best Local Similarity:	98.10%	Mismatches:	0
Query Match:	94.65%	Gaps:	2
DB:	1	Gaps:	1

US-10-089-514-4 (1-103) X AF262220 (1-14159)

QY 1 MetThrGlnAsnGluLeuGln---ArgLeuArgAlaGluLeuAspAlaLeuAspG1 19

Db 11142 ATGACCGAGCAGAACGAGCTGCGGCTGCGCGCGAGCTCGACGCCCTCGACGG 11201

QY 19 YThrLeuLeuAspThrValArgArgArgIleAspLeuGlyValArgIleAlaArgTyrLys 39

Db 11202 GACGCTTCTGGACACGGTGCGGCGCCGATCGACCTCGGTGTCCGCATCGCGCGGTACAA 11261

39 sSerArghIsgIyValPrometmetGlnProGlyArgValSerLeuValLysAspArgAl 59

Db 11262 GTC CCGG CACGGCGTCCCGATGATGCAGCCCGGGGTCAAGGACAGGGC 11321

QY 59 aaIaArGTyRAIaaIaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyraSpva 799

Db 11322 CGCCGCTACGCCGCCGACCAACGGCCTCGACGAATCGTTCCTGTGAACCTCTACGACGT 11381

QY 79 I I I I I e T h r G l u m e t C y s a r g v a l G l u a s p l e u a l m e t S e r a r g g l u s e r l e u t h r a l 99

Db 11382 GATCATCAGGAGATGTGCCGCTCGAGACCTGGTGATGAGCCGGAGAGACCTGACGGC 11441

QY 99 agluaspargarg 103

Db 11442 CGAGGACCGCG 11454

## RESULT 5

A48325 387 bp

ACCESSION	AA8325
DEFINITION	sequence 3 from patent WO9601901.

VERSION A48325.1 GI:2302118

**SOURCE**

1000

## REFERENCE

## AUTHORS

JOURNAL

COMMENT

100

**Source**

ОБЪЕКТЫ

Pred. No.:

Score:	256.00	Matches:	51
Percent Similarity:	71.08%	Conservative:	8
Best Local Similarity:	61.45%	Mismatches:	24
Query Match:	49.81%	Indels:	0
GB:	6	Gaps:	0

US-10-089-514-4 (1-103) X A48325 (1-387)



QY	7	LeuGlnAArgLeuAArgAlGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrValAArg	26
Db	64	CTCGACGGCGCTCGCGCGCGCGCTGGAACCGCGGACGCCGCCCTCTGAGAGCCGCTCCGC	123
QY	27	ArgArgIleAspLeuGlyValArgIleAlaArgTyrIlySerArgHisGlyValAProMet	46
Db	124	ACACGGCTGACATCTGCTCGCGGATCGGCGGAGTCAAGCGCGCTCCACAGGTGCGATG	183
QY	47	MetGlnProGlyAArgValSerLeuValIysAspArgAlaAlaArgTyrAlaAlaAspHis	66
Db	184	ATGCAGGCCCAACCGGATGCCAGGTCCACAGCCCAACGCCGCCGCTACGCCGCGACAC	243
QY	67	GlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIleThrGluMetCysArg	86
Db	244	GGCATCGACCCCGCTTCTCTGCGACCTGTACAGACAGATCATACCGAGACTGTCGCC	303
QY	87	ValGluAsp	89
Db	304	CTCGAGGAC	312
RESULT 6			
LOCUS	AR198355	387 bp	DNA
DEFINITION	Sequence 4 from patent US 6352839.		linear
ACCESSION	AR198355		
VERSION	AR198355.1		GI:20248204
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 387)		
AUTHORS	Blanc,V., Thibaud,D., Bamas-Jacques,N., Blanche,F., Crouzet,J., Barriere,J.-C., Debussche,L., Famechon,A., Paris,J.-M. and Dutric-Roesel,G. Streptogramins for preparing same by mutasynthesis Patent: US 6352839-A 4 05-MAR-2002; location/Qualifiers		
TITLE	1. 387		
JOURNAL	/organism="unknown"		
FEATURES	/mol_type="unassigned DNA"		
Source			
ORIGIN			
Alignment Scores:			
Pred. No.:	9.32e-19	Length:	387
Score:	256.00	Matches:	51
Percent Similarity:	71.08%	Conservative:	8
Best Local Similarity:	61.45%	Mismatches:	24
Query Match:	49.81%	Indels:	0
DB:	6	Gaps:	0
US-10-089-514-4 (1-103) x AR198355 (1-387)			
QY	7	LeuGlnAArgLeuAArgAlGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrValAArg	26
Db	64	CTCGACGGCGCTCGCGCGCGCGCTGGAACCGCGGACGCCGCCCTCTGAGAGCCGCTCCGC	123
QY	27	ArgArgIleAspLeuGlyValArgIleAlaArgTyrIlySerArgHisGlyValAProMet	46
Db	124	ACACGGCTGACATCTGCTCGCGGATCGGCGGAGTCAAGCGCGCTCCACAGGTGCGATG	183
QY	47	MetGlnProGlyAArgValSerLeuValIysAspArgAlaAlaArgTyrAlaAlaAspHis	66
Db	184	ATGCAGGCCCAACCGGATGCCAGGTCCACAGCCCAACGCCGCCGCTACGCCGCGACAC	243
QY	67	GlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIleThrGluMetCysArg	86
Db	244	GGCATCGACCCCGCTTCTCTGCGACCTGTACAGACAGATCATACCGAGACTGTCGCC	303
QY	87	ValGluAsp	89
Db	304	CTCGAGGAC	312
RESULT 7			

LOCUS	2888 bp	DNA	linear	PAT 07-MAR-1997
DEFINITION	Sequence 1 from Patent WO9601901.			
ACCESSION	A48323			
VERSION	A48323.1	GI:2302116		
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS	Streptomyces pristinaespiralis Streptomyces pristinaespiralis Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces. 1 (bases 1 to 2888) Blanc, V., Thibaut, D., Bamas-Jacques, N., Blanche, F., Crouzet, J., Barriere, J., Debussche, L., Famechon, A., Paris, J. and Dutric-Roseet, G. STREPTOMYCIANES AND METHOD FOR PREPARING SAME BY MUTASYNTHESIS Patent: WO 9601901-A 1 25-JAN-1996; RHONE-POULENC RORER SA (FR) Other publication AU 2891295 960209 Other publication FR 2722210 960112.			
FEATURES				
source	1..2888 /organism="Streptomyces pristinaespiralis" /mol_type="unassigned DNA" /db_xref="taxon:38300"			
ORIGIN				
Alignment Scores:				
Score, No.:	7.85e-18	Length:	2888	
Percent Similarity:	256.00	Matches:	51	
Best Local Similarity:	71.08%	Conservative:	8	
Query Match:	61.45%	Mismatches:	24	
DB:	49.81%	Indels:	0	
	6	Gaps:	0	
US-10-089-514-4 (1-103) x A48323 (1-2888)				
Qy	7	LeuGlnArgLeuAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrValArg	26	
Db	1936	CTTCACCGCGGTGCGCGCCGCTGAGACCGCGGACCGCCGCTGCGAGCGCGTCCG	1995	
Qy	27	ArgArgIleAspLeuGlyValArgIleAlaArgTyrTyrSerArgHisGlyValProMet	46	
Db	1996	ACACGCTGAGCATCTCCCTGCGCATCGGGAAGTACAGCGCCCTCCACCAAGTCCGATG	2055	
Qy	47	MetGlnProGlyArgValSerLeuValIleAspArgIleAlaArgTyrAlaIleAspHis	66	
Db	2056	ATGACGCGCCACCGGATCGCCGAGTCCACGACCGCCGCTACGCGCGGACAC	2115	
Qy	67	GlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIleThrGluMetCysArg	86	
Db	2116	GGCATCCACCCCGCTTCTGCGCACCGCTGACACAGATCATCACCGAGACCTGCCG	2175	
Qy	87	ValGluAsp	89	
Db	2176	CTCGAGGAC	2184	
RESULT 8				
LOCUS	ARI98353	2888 bp	DNA	linear
DEFINITION	Sequence 1 from patent US 6352839.			PAT 20-APR-2002
ACCESSION	ARI98353			
VERSION	ARI98353.1	GI:20248202		
KEYWORDS				
SOURCE				
ORGANISM	Unknown.			
REFERENCE	Unclassified.			
AUTHORS	1 (bases 1 to 2888) Blanc, V., Thibaut, D., Bamas-Jacques, N., Blanche, F., Crouzet, J., Barriere, J.-C., Debussche, L., Famechon, A., Paris, J.-M. and Dutric-Roseet, G. Streptogramins for preparing same by mutasynthesis Patent: US 6352839-A 1 05-MAR-2002; Location/Qualifiers			
TITLE				
JOURNAL				
FEATURES				



	source	1..2888 /organism="unknown" /mol_type="unassigned DNA"
	Alignment Scores:	
Pred. NO.:	7,85e-18	Length: 2888
Score:	256.00	Matches: 51
Percent Similarity:	71.08%	Conservative: 8
Best Local Similarity:	61.45%	Mismatches: 24
Query Match:	49.81%	Indels: 0
DB:	6	Gaps: 0
US-10-089-514-4 (1-103) x ARI198353 (1-2888)		
Oy	7 LeuglnArgLeuAArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrValArg 26	
Db	1936 CTCACAGCGCTGCAGCGCCGCCGTGGACGCCCGCAGCGCCGCTGTCGATCACGCCGTCGCG 1995	
Oy	27 ArgArggllleAspleuGlyValArglIleAlaArgTyrIlysSerAArgHlglyValPromet 46	
Db	1996 ACAGCGCTCGAGCATCTGCCTCGCATCGGACGAGTACAGACGCCCTCCACCAGSTGCCGATG 2055	
Oy	47 MetGlnProGlyAArgValSerIeuValIysAsprrglAlaAlaArgTyrAlaAlaSPHis 66	
Db	2056 ATGCAGGCCCAACCGGATGCCAGGTCCAGGCCAACGCCGCCGCCGCTACGCCGCCGACAC 2115	
Oy	67 GlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIlelleIrngIueTcysArg 86	
Db	2116 GGCATCGACCCCGCCTCTCCTCGCGACCCCTGTACGACGATCATCACGAGACTGCCGCG 2175	
Oy	87 ValGIuAsp 89	
Db	2176 CTCGAGGAC 2184	
RESULT 9		
SPU60417	4740 bp DNA linear BCT 07-MAR-1997	
LOCUS	Streptomyces pristinaeaeSPIralls 4-dimethylamino-L-phenylalanine precursor biosynthesis (papA, papC, papB, papM) genes, complete cds.	
DEFINITION		
ACCESSION	U60417	
VERSION	U60417.1 GI:1575335	
KEYWORDS		
SOURCE	Streptomyces pristinaeaeSPIralls	
ORGANISM	Streptomyces pristinaeaeSPIralls	
REFERENCE	Bacteria; Actinobacteriae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.	
AUTHORS	1 (bases 1 to 4740) Blanc,V., Gil,P., Bamas-Jacques,N., Lorenzon,S., Zagorec,M., Schleuniger,J., Bisch,D., Blanche,F., Debussche,L., Crouzet,J. and Thibaut,D.	
TITLE	Identification and analysis of genes from Streptomyces pristinaeaeSPIralls encoding enzymes involved in the biosynthesis of the 4-dimethylamino-L-phenylalanine precursor of pristinamycin I	
JOURNAL	Mol. Microbiol. 23 (2), 191-202 (1997)	
MEDLINE	97197164	
PUBMED	9044253	
REFERENCE	2 (bases 1 to 4740) Blanc,V., Gil,P., Bamas-Jacques,N., Lorenzon,S., Schleuniger,J., Bisch,D., Blanche,F., Debussche,L., Crouzet,J. and Thibaut,D.	
AUTHORS	Direct Submission Submitted (11-JUN-1996) Recherche Pharmaceutique, Rhone-Poulenc Rorer, 13 quai Jules Guesde, B. P. 14, Vitry sur Seine cedex 94403, France	
TITLE		
JOURNAL		
FEATURES		
source	Location/Qualifiers	
	1..4740	/organism="Streptomyces pristinaeaeSPIralls"
		/mol_type="Genomic DNA"
gene		/db_xref="taxon:38300"
	68..2227	
CDS		/gene="papA"
	68..2227	

[illegible]



DB:	1	Gaps:	0
US-10-089-514-4 (1-103) × SP060417 (1-4740)			
OY	7	LeuGlnAArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrValArg	26
DB	3476	CTCGACGGCGCTGCGGCCCGCTGGAACCGCGGACGCCGCCCTGCTGAGACCGCTCCG	3535
OY	27	ArgArgIleAspLeuGlyValArgIleAlaArgIlyTrpLysSerArgHisGlyValProMet	46
DB	3536	ACACCGCTCGACATCTGCGCTGCGCATCGGCGAGTCAAGACGCCCTCCACCAAGTGGCGATG	3595
OY	47	MetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyrAlaAlaAspHis	66
DB	3596	ATGCAAGCCCAACCGATAGCCCAAGTCCACGCGCAACGCCGCCGCTAGCGCGCGCAAC	3655
OY	67	GlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIleThrGluMetCysArg	86
DB	3656	GGCATCGACCCGCGCTTCTGCGCACCTTGACGACAGATCATATACCGAAGCTGCGCG	3715
OY	87	ValGluAsp	89
DB	3716	CTCGAGGAC	3724
RESULT 10			
LOCUS	BX571871	349944 bp	DNA
DEFINITION	Phototrichobabds luminescens subsp. laumondii T701 complete genome;		
ACCESSION	BX571871	BX470251	
VERSION	BX571871.1	GI:36786846	
KEYWORDS	complete genome.		
SOURCE	Phototrichobabds luminescens subsp. laumondii T701		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
REFERENCE	1	Duchaud, E., Rusniok, C., Frangeul, L., Buchrieser, C., Taourit, S.,	
AUTHORS	Bocs, S., Bouraux-Eude, C., Chandelier, M., Dassa, E., Decose, R.,		
	Dezelic, S., Freysinet, G., Gaudriault, S., Givaudan, A., Glaeer, P.,		
	Medigue, C., Lanois, A., Powell, K., Siguier, P., Wingate, V.,		
	Zouine, M., Boemare, N., Danchin, A. and Kunst, F.		
	Complete genome sequence of the entomopathogenic bacterium		
	Phototrichobabds luminescens		
	Nat. Biotechnol. 11 (1) (2003) In press		
TITLE	2	Duchaud, E., Frangeul, L., Rusniok, C. and Kunst, F.	
JOURNAL	Direct Submission		
REFERENCE	Submitted (23-Apr-2003) L. Frangeul, Institut Pasteur, Genopole, 25		
AUTHORS	rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail:		
JOURNAL	lfrangeul@pasteur.fr, fkunst@pasteur.fr		
	Location/Qualifiers		
FEATURES	1..349944		
source	/organism="Phototrichobabds luminescens subsp. laumondii T701"		
	/mol_type="Genomic DNA"		
	/strain="T701"		
	/db_xref="taxon:243265"		
	complement(1142..3038)		
	/gene="prpB"		
	/locus_tag="plu3539"		
	complement(1142..3028)		
	/gene="prpE"		
	/locus_tag="plu3539"		
	/codon_start=1		
	/transl_table=11		
	/product="Propionate--CoA ligase (Propionyl-CoA		
	synthetase)"		
	/protein_id="CAE15912.1"		
	/db_xref="GI:36786847"		
	/translat="MSFAAFQSGIDDPNAFMABOAOIRFMQHPFEOVLDSNPPPAR		
	MFQCGDVLVYMPHIALALVYLACARIGVHVSVRGPFASQSLATRLDNEPVLVY		
	LGVKGSDVLYVMPHIALALVYLACARIGVHVSVRGPFASQSLATRLDNEPVLVY		
	SADGGARGCGIIPYKPLUDEAILELANHPRVILVNWGLAMVWVPGDIDFATLRQK		
gene			
CDS			

YEDADI PTWMESENSTSVL YTSCTTGTCRGVQVQDUGVAVLAATL TSDMDI LFGCAAGVAF  
FFTSIDICMTHGSIYLAIVLAALMATHLITBGLTRPRDAGIWKATVEKIOYTRFSPAR  
TAIRLKYALPIECIGKYDI SSLKTLYLAGEPLDRTARMALEAINVIVINDYWTQFGR  
MPTIMAIATTLDRPSRFGSTGFPMYGPHVKLINLTGECAGCDNEKMYLVI EGFLPPRC  
IOTIIGDADTRFINTYWRHFBAVYSTEDWGLRSDGYFLIAGRSDVIYVAGHGLGRR  
ELEBETIAHEBAVFAVIGIKDAIKGOVAFAVLXDGKEIQNAHEFAALEKVLMGVAV  
NKQISVSRPARIYFVSQLPKTRSGKMLRTMQAICBGRBGDSLLENPASLDVIRE  
AVPC"

RBS  
/gene="prpB"  
/locus\_tag="plu3539"  
complement (3101. .4565)  
/gene="prpB"  
/locus\_tag="plu3540"  
complement (3101. .4552)  
/locus\_tag="plu3540"  
/codon\_start=1  
/transl\_table=11  
/product="2-methylcitrate dehydratase"  
/protein\_id="CAE1591.1"  
/db\_xref="gi:36766848"  
/translation="MSTSVASNHEBYDOVLADIVYMEVALITSSVAEYAYVCELDPT  
LGGCEBALEYACKKLLGPIYPTGVTVNNGRVPETQQLDPVQAAPIRIGMIRLDPN  
DTWLAELGHSNDLGGTLAVADWLSNAVAGGKPLTIDVLTGMVKAHEIQCLAL  
ENAFKQVLDLHVVLVKASTAVVAMGLNREBILSAVSLVADWGSILRTYRANPVG  
SRKSAKADLSRAVRLALMAQKEMGYPSTLTAKTGWFPYDVFNGQPFPORAYGV  
VMENLTFLIAPAEFHAQTAVEAIALSHOOLKDGAKTINDITKTIITHCACRIIDK  
OSPNNPNDPRHCIOYMAVPLIRGRITADENSVADRIALREKIVICIELDPT  
RDHYDPERKSIANMLTYTFTDGSQLDYKVEFPIGHARRKEGIPLLKKFRTNLAQ  
FPEIQRRILSVSLDCSLBESMPVNEITLDLTVI"  
complement (4560. .4565)

RBS  
/gene="prpB"  
/locus\_tag="plu3540"  
complement (4607. .5787)  
/gene="prpC"  
/locus\_tag="plu3541"  
complement (4607. .5773)  
/gene="prpC"  
/locus\_tag="plu3541"  
/codon\_start=1  
/transl\_table=11  
/product="2-methylcitrate synthase (Methylcitrate  
synthase) (Citrate synthase 2)"  
/protein\_id="CAE1591.1"  
/db\_xref="gi:36766849"  
/translation="MSBONTTELTTPKPKSVALSQSVSNALCTVGTGNDLHR  
GYLDLDTQCEFEVAVLIVHEKLPRAEILAAKALTLNRGPSVKALETLPLPAA  
AHPDMVMTQGVSLGCTLPBKQDNLNLAGADDIADRLASLSLILVYVHSHNGRIE  
VETDVSIGGFHLHLHGKTCSEKAMHTSLIYAEHFSNASTFTSRVIACTGTDI  
YSATIGALIGLRGPHGAGNVSPEIORYENPDEARDIARLARKEVILGCHPQVY  
TISDPRHVITDVALQLSKEAGTKMENIDRLAEVAMWAKMKPNPNDWFSAYSNHM  
GVPLDMFPLVAVIARTGMAAHIIEQIDNKIIRPSADYTGPEKRVPTPIGKG"

RBS  
/gene="prpB"  
/locus\_tag="plu3541"  
complement (5834. .6736)  
/gene="prpB"  
/locus\_tag="plu3542"  
complement (5834. .6724)  
/gene="prpB"  
/locus\_tag="plu3542"  
/codon\_start=1  
/transl\_table=11  
/product="Probable methylisocitrate lyase  
(2-methylisocitrate lyase)"  
/protein\_id="CAE1591.1"  
/db\_xref="gi:36766850"

gene  
/translation="MTLSAGLAFFRAIDAESPLOVGVGTINANHALLAKRAGKATYLR  
SGGVSAGSLGMPDGLSTLDVLTIDHRTIDCDLPLVADNIGSSSAFVAATYLR  
SMIRKAGAGHIIEOVAKSGRPNKEIYSKEMDRICAADAPEADHGFITMGATP  
ALAYGPAALDRKATYENAGDMLPEALITELHMKAFASITGCVPLANVITIGCATP  
LTFTEELKSAVSLTALYPLSAFRAMNMAAOVYITLRBCTQSVISMMQTRNELVES

CDS  
/gene="prpB"  
/locus\_tag="plu3540"  
complement (3101. .4565)  
/gene="prpB"  
/locus\_tag="plu3540"  
/codon\_start=1  
/transl\_table=11  
/product="2-methylcitrate dehydratase"  
/protein\_id="CAE1591.1"  
/db\_xref="gi:36766848"  
/translation="MSTSVASNHEBYDOVLADIVYMEVALITSSVAEYAYVCELDPT  
LGGCEBALEYACKKLLGPIYPTGVTVNNGRVPETQQLDPVQAAPIRIGMIRLDPN  
DTWLAELGHSNDLGGTLAVADWLSNAVAGGKPLTIDVLTGMVKAHEIQCLAL  
ENAFKQVLDLHVVLVKASTAVVAMGLNREBILSAVSLVADWGSILRTYRANPVG  
SRKSAKADLSRAVRLALMAQKEMGYPSTLTAKTGWFPYDVFNGQPFPORAYGV  
VMENLTFLIAPAEFHAQTAVEAIALSHOOLKDGAKTINDITKTIITHCACRIIDK  
OSPNNPNDPRHCIOYMAVPLIRGRITADENSVADRIALREKIVICIELDPT  
RDHYDPERKSIANMLTYTFTDGSQLDYKVEFPIGHARRKEGIPLLKKFRTNLAQ  
FPEIQRRILSVSLDCSLBESMPVNEITLDLTVI"  
complement (4560. .4565)

RBS  
/gene="prpB"  
/locus\_tag="plu3540"  
complement (4607. .5787)  
/gene="prpC"  
/locus\_tag="plu3541"  
complement (4607. .5773)  
/gene="prpC"  
/locus\_tag="plu3541"  
/codon\_start=1  
/transl\_table=11  
/product="2-methylcitrate synthase (Methylcitrate  
synthase) (Citrate synthase 2)"  
/protein\_id="CAE1591.1"  
/db\_xref="gi:36766849"  
/translation="MSBONTTELTTPKPKSVALSQSVSNALCTVGTGNDLHR  
GYLDLDTQCEFEVAVLIVHEKLPRAEILAAKALTLNRGPSVKALETLPLPAA  
AHPDMVMTQGVSLGCTLPBKQDNLNLAGADDIADRLASLSLILVYVHSHNGRIE  
VETDVSIGGFHLHLHGKTCSEKAMHTSLIYAEHFSNASTFTSRVIACTGTDI  
YSATIGALIGLRGPHGAGNVSPEIORYENPDEARDIARLARKEVILGCHPQVY  
TISDPRHVITDVALQLSKEAGTKMENIDRLAEVAMWAKMKPNPNDWFSAYSNHM  
GVPLDMFPLVAVIARTGMAAHIIEQIDNKIIRPSADYTGPEKRVPTPIGKG"

RBS  
/gene="prpB"  
/locus\_tag="plu3541"  
complement (5834. .6736)  
/gene="prpB"  
/locus\_tag="plu3542"  
complement (5834. .6724)  
/gene="prpB"  
/locus\_tag="plu3542"  
/codon\_start=1  
/transl\_table=11  
/product="Probable methylisocitrate lyase  
(2-methylisocitrate lyase)"  
/protein\_id="CAE1591.1"  
/db\_xref="gi:36766850"

gene  
/translation="MTLSAGLAFFRAIDAESPLOVGVGTINANHALLAKRAGKATYLR  
SGGVSAGSLGMPDGLSTLDVLTIDHRTIDCDLPLVADNIGSSSAFVAATYLR  
SMIRKAGAGHIIEOVAKSGRPNKEIYSKEMDRICAADAPEADHGFITMGATP  
ALAYGPAALDRKATYENAGDMLPEALITELHMKAFASITGCVPLANVITIGCATP  
LTFTEELKSAVSLTALYPLSAFRAMNMAAOVYITLRBCTQSVISMMQTRNELVES

CDS  
/gene="prpB"  
/locus\_tag="plu3540"  
complement (3101. .4565)  
/gene="prpB"  
/locus\_tag="plu3540"  
/codon\_start=1  
/transl\_table=11  
/product="2-methylcitrate dehydratase"  
/protein\_id="CAE1591.1"  
/db\_xref="gi:36766848"  
/translation="MSTSVASNHEBYDOVLADIVYMEVALITSSVAEYAYVCELDPT  
LGGCEBALEYACKKLLGPIYPTGVTVNNGRVPETQQLDPVQAAPIRIGMIRLDPN  
DTWLAELGHSNDLGGTLAVADWLSNAVAGGKPLTIDVLTGMVKAHEIQCLAL  
ENAFKQVLDLHVVLVKASTAVVAMGLNREBILSAVSLVADWGSILRTYRANPVG  
SRKSAKADLSRAVRLALMAQKEMGYPSTLTAKTGWFPYDVFNGQPFPORAYGV  
VMENLTFLIAPAEFHAQTAVEAIALSHOOLKDGAKTINDITKTIITHCACRIIDK  
OSPNNPNDPRHCIOYMAVPLIRGRITADENSVADRIALREKIVICIELDPT  
RDHYDPERKSIANMLTYTFTDGSQLDYKVEFPIGHARRKEGIPLLKKFRTNLAQ  
FPEIQRRILSVSLDCSLBESMPVNEITLDLTVI"  
complement (4560. .4565)

RBS  
/gene="prpB"  
/locus\_tag="plu3540"  
complement (4607. .5787)  
/gene="prpC"  
/locus\_tag="plu3541"  
complement (4607. .5773)  
/gene="prpC"  
/locus\_tag="plu3541"  
/codon\_start=1  
/transl\_table=11  
/product="2-methylcitrate synthase (Methylcitrate  
synthase) (Citrate synthase 2)"  
/protein\_id="CAE1591.1"  
/db\_xref="gi:36766849"  
/translation="MSBONTTELTTPKPKSVALSQSVSNALCTVGTGNDLHR  
GYLDLDTQCEFEVAVLIVHEKLPRAEILAAKALTLNRGPSVKALETLPLPAA  
AHPDMVMTQGVSLGCTLPBKQDNLNLAGADDIADRLASLSLILVYVHSHNGRIE  
VETDVSIGGFHLHLHGKTCSEKAMHTSLIYAEHFSNASTFTSRVIACTGTDI  
YSATIGALIGLRGPHGAGNVSPEIORYENPDEARDIARLARKEVILGCHPQVY  
TISDPRHVITDVALQLSKEAGTKMENIDRLAEVAMWAKMKPNPNDWFSAYSNHM  
GVPLDMFPLVAVIARTGMAAHIIEQIDNKIIRPSADYTGPEKRVPTPIGKG"

RBS  
/gene="prpB"  
/locus\_tag="plu3541"  
complement (5834. .6736)  
/gene="prpB"  
/locus\_tag="plu3542"  
complement (5834. .6724)  
/gene="prpB"  
/locus\_tag="plu3542"  
/codon\_start=1  
/transl\_table=11  
/product="Probable methylisocitrate lyase  
(2-methylisocitrate lyase)"  
/protein\_id="CAE1591.1"  
/db\_xref="gi:36766850"

gene  
/translation="MTLSAGLAFFRAIDAESPLOVGVGTINANHALLAKRAGKATYLR  
SGGVSAGSLGMPDGLSTLDVLTIDHRTIDCDLPLVADNIGSSSAFVAATYLR  
SMIRKAGAGHIIEOVAKSGRPNKEIYSKEMDRICAADAPEADHGFITMGATP  
ALAYGPAALDRKATYENAGDMLPEALITELHMKAFASITGCVPLANVITIGCATP  
LTFTEELKSAVSLTALYPLSAFRAMNMAAOVYITLRBCTQSVISMMQTRNELVES

CDS  
/gene="prpB"  
/locus\_tag="plu3540"  
complement (3101. .4565)  
/gene="prpB"  
/locus\_tag="plu3540"  
/codon\_start=1  
/transl\_table=11  
/product="2-methylcitrate dehydratase"  
/protein\_id="CAE1591.1"  
/db\_xref="gi:36766848"  
/translation="MSTSVASNHEBYDOVLADIVYMEVALITSSVAEYAYVCELDPT  
LGGCEBALEYACKKLLGPIYPTGVTVNNGRVPETQQLDPVQAAPIRIGMIRLDPN  
DTWLAELGHSNDLGGTLAVADWLSNAVAGGKPLTIDVLTGMVKAHEIQCLAL  
ENAFKQVLDLHVVLVKASTAVVAMGLNREBILSAVSL



RBS  
INYYDFEOKLIDLFQKSOQ"  
complement (6731.. 6736)  
/gene="prpb"  
/locus\_tag="plu3542"  
7047.. 8724  
/locus\_tag="prpr"  
/locus\_tag="plu3543"  
7047.. 7052  
/gene="prpr"  
/locus\_tag="plu3543"  
7057.. 8724  
/gene="prpr"  
/locus\_tag="plu3543"  
/codon\_start=1  
/transl\_table=1  
/product="Propionate catabolism operon regulatory protein"  
/protein\_id="CAE15916.1"  
/db\_xref="GI:36786851"  
/translation="MAETPSAERDNRDKPVITWTVSARLPDLPRDITPEFHQNIPT  
IOLGSRNVOHQRORLTTECCDAVLCAGSNAYAKSLISVPVITAKASGDFMOLAR  
ARQISRTIGVITVQNTLPALBEPQDFRLEQNSVYTEEDARQVNAKAIIGIKTV  
GAGLTIDAELELVGVFVSADSIROAFHDALDIARTTLLSOSISAYPTENNLRV  
HTINDIRDSAVMEHVRNTIMLVARSPTVLIQESGKELVQALIHRYALRYCGV  
AGKHQRPVAVNCAIASLELAEELFGEBAFPGSGRRGAGLEVAHGGTLELDI  
GEMPIPLQITPLRLVLEEKSVVRGVGHPRIANDVRLICATHCDLDTWODRPRDLY  
RLSVARTINIPALREKHDIGTILAEYLRARPAALNLYSOSRVEIATCEBALDIYH  
PENVEKLEMERLTLVYSAPEHILTEPELLTASPERQSLIKQAEQSMKSPSSIQQ  
SLVPLVSHVELTKRPSGNHHAETVEYLGISRTTLMRLKEIAREE"  
complement (8738.. 9709)  
/locus\_tag="plu3544"  
complement (8738.. 9709)  
/locus\_tag="plu3544"  
/function="Unknown"  
/notes="unnamed protein product; similar to probable  
transcription regulator"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAE15917.1"  
/db\_xref="GI:36786852"  
/translation="MNITQIFFLFPNVNLDLSGLQVPTTANELAEKGRHRRPVAI  
KVADDPQWTOAGITVQABPLPLDLSPADTLIVAGIGVESANHPPLRNMLKQHA  
SLARIVVSGSAFLIACGLDGRVVTWSSCAELARRSKLRVCDPIFDQDGV  
WTSAGVTGIDILALAEEDRGKLDLVDADIDYMLKRPQGGQCPAPLALOSSTK  
PADLHAMTNNIAGDLSVATTLATPANSSENSFVTHYCHAGITPAKAVEPRULEARH  
LVESITLSTKICIASCGFCEDTLRRSFFRAFSSTPDYRKHFKTTP"  
complement (10784  
/locus\_tag="plu3545"  
10003.. 10008  
/locus\_tag="plu3545"  
10017.. 10784  
/locus\_tag="plu3545"  
/function="Unknown"  
/notes="unnamed protein product; Some similarities with  
unknown protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAE15918.1"  
/db\_xref="GI:36786853"  
/translation="MKNTSKHEPALSLINASSGKPKTLAFVIFPGMTLDMIGPAT  
VLGSGMNITNYVRDILPYRTELAHISFSSTTPEQOKNTDILCVGCTGNPVALIAE  
LDMELMTHKVGSKASWTSVCTGSLILGACGLNGYQALATHWSNITDLUNKFDPIPVK  
KVIDIRQAVTGGGVTAGIDFGLTLLAIKGEVYAQAVQALMAYSPBPENSGMPEFAD  
KEITATATQAIFSNVEKQTPYKEILIKSAKRISDIY"  
complement (1153.. 12540)  
/gene="ISPlu15A"  
/locus\_tag="plu3546"  
complement (11533.. 12540)  
/gene="ISPlu15A"  
/locus\_tag="plu3546"  
/function="Transposase"  
/codon\_start=1  
/transl\_table=1  
/product="Transposase, ISNCY family"

US-10-089-514-4 (1-103) x BX571871 (1-349944)	Alignment Scores:	Alignment Scores:
OY 1 MetHrcIuGlnaAngIuLeuGlnArgLeuArgAlaGluLeuAphAlaLeuAapGlyThr 20	Score: 120.00	Length: 349944
Db 32716 ATAAATATAACGGAATATTACCACTTACACGGGAATCTTTGGATTATTAATATGCGAG 3265	Percent Similarity: 55.06%	Matches: 25
OY 21 LeuLeuAapSprHrValArgArgArgIleAapLeuGlyValArgIleAlaArgIlyrLysSer 40	Best Local Similarity: 28.09%	Mismatches: 40
Db 32656 ATTTTGGCAGCTGCTTCTGCAACGATATAAGTTGTATGAAATCGCTGAGCGTAAAGCG 3259	Query Match: 23.35%	Indels: 0
OY 41 ArgHisGlyValPProMetMetGlnProGlyArgValSerLeuValLysAapArgAlaAla 60	DB: 1	Gaps: 0
Db 32596 GGGCAGAGATATCCGATGATGACGCCACAACGATATATCTCGTATGATGATGTTGCCG 3253		
OY 61 ArgIlyrAlaAlaAapHisGlyLeuAapGluSerPheLeuValLeuLeuIlyrAapValIle 80		
Db 32536 GATTAATCAACCGATTTTGGGTTTACCGGCTGAATATATCTGAATCAATATTCAACGTGT 3247		
OY 81 IleHrcIuMetCysArgValGluAap 89		
Db 32476 ATTGAAGACACTTGCCTCGAGAAAGA 32450		
RESULT 11		
LOCUS AX770909 349980 bp DNA linear PAT 02-JUL-2003		
DEFINITION Sequence 40 from Patent WO02094867.		
ACCESSION AX770909		
VERSION AX770909.1 GI:32438073		
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
FEATURES		
source		
ORIGIN		



DB:	6	Gaps:	0
US-10-089-514-4 (1-103) x AX770909 (1-349860)			
OY	1	MecthrGlUGlnAsnGluLeuGlnaGLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20	
Db	113584	ATTAATTAACACGAAATATATTACACCTTACAGGAACTCTTGATGTATTAATATAGCAG 113643	
OY	21	LeuLeuAspThrValaArgArgArgIleAspLeuGlyValaArgIleAlaArgTyrLysSer 40	
Db	113644	ATTTTGGCAGCTCTCTTCTGAACGTATGAAGAGTTGTATGAAATCCCTGAGTAAAGCG 113703	
OY	41	ArgHisIsgIValaPromeMetGlnPProGlyArgValSerLeuValIlysAspAlaAla 60	
Db	113704	GGCGACGATATTCCTCGATGTAGAGCCACACCGATTAATCTCGTATGATGATGTTGCCG 113763	
OY	61	ArgTyrAlaAlaAspHisIsgILeuAspGluSerPheLeuValaAsnLeuTyrAspValIle 80	
Db	113764	GATTAATCAACCGATTTTGGGTTCACGCCCTGATATATCATGATCAATATTTCACATGGTT 113823	
OY	81	IleThrGluMetCysArgValGluAsp 89	
Db	113824	ATTGAAGAGACTTGGCCGTCGAGAGAA 113850	
RESULT 12			
LOCUS	AE017175	300454 bp	DNA linear BCT 02-SEP-2003
DEFINITION	Porphyromonas gingivalis W83 section 4 of 8 of the complete genome.		
ACCESSION	AE017175	AE015924	
VERSION	AE017175.1	GI:34396931	
KEYWORDS			
SOURCE			
ORGANISM	Porphyromonas gingivalis W83 Porphyromonas gingivalis W83 Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales; Porphyromonadaceae; Porphyromonas.		
REFERENCE	1 (bases 1 to 300454)		
AUTHORS	Nelson,K., Fleischmann,R., DeBoy,R., Paulsen,I., Fouts,D., Eisen,J., Daugherty,S., Dodson,R., Durkin,A., Gwim,M., Haft,D., Kolonay,J., Nelson,M., White,O., Mason,T., Tallon,L., Gray,J., Granger,D., Tetcelin,H., Dong,H., Galvin,J., Duncan,M., Dewhirst,F. and Fraser,C.		
TITLE	Complete Genome Sequence of the Oral Pathogenic Bacterium		
JOURNAL	Porphyromonas gingivalis Strain W83		
PUBMED	J. Bacteriol. 185 (18), 5591-5601 (2003)		
REFERENCE	12349112		
AUTHORS	2 (bases 1 to 300454)		
	Nelson,K., Fleischmann,R., DeBoy,R., Paulsen,I., Fouts,D., Eisen,J., Daugherty,S., Dodson,R., Durkin,A., Gwim,M., Haft,D., Kolonay,J., Nelson,M., White,O., Mason,T., Tallon,L., Gray,J., Granger,D., Tetcelin,H., Dong,H., Galvin,J., Duncan,M., Dewhirst,F. and Fraser,C.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-OCT-2002) The Institute for Genomic Research, 9712		
FEATURES	Medical Center Dr., Rockville, MD 20850, USA		
SOURCE	Location/Qualifiers		
	1..300454		
	/organism="Porphyromonas gingivalis W83"		
	/mol_type="genomic DNA"		
	/strain="W83"		
	/db_xref="taxon:242619"		
	177..1757		
	/locus_tag="PG0847"		
	177..1757		
	/locus_tag="PG0847"		
	/codon_start=1		
	/transl_table=11		
	/product="conserved domain protein"		
	/protein_id="AAO65996.1"		
	/db_xref="GI:34396932"		
	/translation="MTDYAHKMSLDLRMTDLERSPHIDNKASHKAKMLNDYYQKDA AKNQFVSDICTYKPDGEFNVAYSKTKRLVEDYHNPSEHVPFIOCKKBEKKAKMVD AKNRDDITIPGSTMLGTGAQQAVALIHLDLPWPSDLEQRNGRAVRKGNMTAK EFADNRKVDVITIAVERSLDSYKFNLLHNKQLFINDLKTTLQSRITDEGSMEDGGMN		
gene			
CDS			

[illegible]



```

CDS      6807..7292
         /locus_tag="PG0853"
         /codon_start=1
         /transl_table=11
         /product="DNA-binding protein, histone-like family"
         /db_xref="gi:34396937"
         /translation="MIKIKAIERRKAGFEKTSKTITMYPAILHSDVKEEFELIVSDEFT
TVSSADIAVEPRRAKVLIRLQDQKSIDCGDMGYRPSRIYAKGASGVDSADKVTVEL
VDKAVIYTPRKVKTKTALKGVMERAEKRLDVPAASSIKPENNNSGGSSSDNNGHAG
I"
gene     7293..8111
         /locus_tag="PG0854"
         /locus_tag="PG0854"
         /locus_tag="PG0854"
         /note="identified by Glimmer2; putative"
         /codon_start=1
         /transl_table=11
         /product="hypothetical protein"
         /protein_id="AAO66002.1"
         /db_xref="gi:34396938"
         /translation="MYPIKIIYVYKAMLDKYIDKAGVLTREAYGVVKKPVLPKIGEVL
GPLSPICLOAQGMKWILINGEESIRLTSEAIRKIETAPERLTPEEPYVVVAIOOL
IGVSSSEIRAEIYANLIIVSSMDVEKKGLVHPAVIDLKTLTPEATIQFCKRPDI
ETLRDRATIKEEEGGFRTIADHKTLISDEVNFMPDNELAYLONLVCLGIKDCBET
FLVIDNKLLENKLTGLESLRNQYVPEQFSIRSEKSFYOTDFGRMFINTIVE"
CDS      complement(8333..8428)
         /locus_tag="PG0855"
         /complement(8333..8428)
         /locus_tag="PG0855"
         /note="identified by Glimmer2; putative"
         /codon_start=1
         /transl_table=11
         /product="hypothetical protein"
         /protein_id="AAO66003.1"
         /db_xref="gi:34396939"
         /translation="WIFSLHHFALICSEMGPFAVSKRAKTYMEVL"
gene     8613..9518
         /locus_tag="PG0856"
         /locus_tag="PG0856"
         /locus_tag="PG0856"
         /note="identified by Glimmer2; putative"
         /codon_start=1
         /transl_table=11
         /product="hypothetical protein"
         /protein_id="AAO66004.1"
         /db_xref="gi:34396940"
         /translation="WDENLRQLASLLTDLTOPAPINTMLKAPVASKRNDIDLWS
IEHLANGBNENPKRLIDAGKYVDIHRCFOEVGINPMAVNOEKVERELHLPLPH
KSLSIVSEELSTSGRGTHLIDIPIEIWHNRHCINDIORAYOFATVASKOIMWKI
KSLILDYFKLTIDKGSLSFLDIKKETPTMOIIGIYVTSGANTANGATITISANTS
INKDMLPELORLISCIEIMKINSDDYEISEELKTEIERDSSPNRIKRGLQAING
IASGIVSGVINQVSEPLVTSALL"
gene     9742..9966
         /locus_tag="PG0857"
         /locus_tag="PG0857"
         /locus_tag="PG0857"
         /note="identified by match to PFAM protein family HM
PF01381"
         /codon_start=1
         /transl_table=11
         /product="transcriptional regulator, putative"
         /protein_id="AAO66005.1"
         /db_xref="gi:34396941"
         /translation="MQARTIIAQYVKKMKRYKYLTOYDLSKAGVGILRFVRELQGT
TLRLDKAVKVFELFSCSECGVPVRKREGDV"
gene     9963..10298
         /locus_tag="PG0858"
         /locus_tag="PG0858"
         /locus_tag="PG0858"
         /note="similar to GB:D90318, GB:D90319, GB:D90320,
GB:D90321, GB:D90322, GB:D90323, GB:D90324, GB:D90325,
GB:D90326, GB:D90327, GB:D90328, GB:D90329, GB:D90330,"

```

Alignment Scores:	
Pred. No.:	2 39
Score:	11.05
Percent Similarity:	54.85
Best Local Similarity:	29.55
Query Match:	22.18
DB:	1
	Gaps: 0
	Length: 300454
	Matches: 26
	Conservative: 22
	Mismatch: 40
	Indels: 0

US-10-089-514-4 (1-103) X AE017175 (1-300454)

QY 3 GlnGlnAsnGluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeu 22

Db 40966 CAAGACGAGAGCTGATCTCTTGGCCATGCAGATTGATCAGATAGATGAGAGTATAGTG 41025

QY 23 AsprThrValArgArgArgIleAspLeuGlyValArgIleAlaArgTyrLysSerArgHis 42

41026 GAATTGCTAGCTCGGCGGATGCAAGTGGCATACGAGATAGGTTTGTTCAAAAAAGAGCAC 41085

43 GlyvalPrometmetGlnProGlyargvalserleuvallylsaspargalaalaarglyr 62

41068 AACCTGGCTGAGTCAAGAACCTCCGTAAGAACAACTACAGCTCAACCGTGTCCGTAAC 41143

[illegible]

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

[illegible]

U67480/c

**DEFINITION** *Methanocaldococcus jannaschii* DSM 2661 section 22 of 150 of the

ACCESSION U67480 L77117

KEYWORDS  
.  
Motivation; Adolescents; Family; Peer; Self

ORGANISM Methanocaldococcus jannaschii DSM 2661  
Archaea: Euryarchaeota: Methanococci: Methanococcales:

REFERENCE 1 (pages 1 to 10632)  
Mecynotocariaceae; Mecynotocarius.

Sutton.G.G., Blake.T.A., Fitzgerald.I.M., Clayton.B.A.,

Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstein, K.G.,

Fuhrmann, J.L., Nguyen, D.T., Utterback, T., Kelley, J.M.,

Roberts, K.M., Kaine, B.B., Borodovsky, M., Klenk, H.P., Fraser, C.M.,

**TITLE** Complete genome sequence of the methanogenic archaeon,

**JOURNAL** Science 273 (5278), 1058-1073 (1996)

PUBMED 8688087

**AUTHORS** . Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D.,

Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. -F.,

Merrick, J. M., Glodok, A., Scott, J. D., Geognagen, N. S., Weidman, J. F.,  
 Puberman, I. I., Newman D. T., Hitterback T., Kelley J. M.

PETERSON, J. D., SADOW, F. W., HANNA, M. C., COLTON, M. D., HURST, M. A.,







/note="similar to SP:P00211 percent identity: 37.04;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="ferredoxin II"  
/protein\_id="AB98238.1"  
/db\_xref="GI:1499030"  
/translation="MENLEKIEELKKIREFLINLEIKKMOELNVDSDIYEAYEKV  
TKIVRPNIKLYQYDAIKEMFYEEYGGKKKDIISWPKIDYRCKNCKCSFCPRG  
VYDAENGKVVKKPKYSCIVNCNCSIMCCENNAIIFPDEKIPRRN"  
6715..7419  
/locus\_tag="MJ0252"  
6715..7419  
/locus\_tag="MJ0252"  
/note="similar to SP:P42405 PID:710637 PID:1805418  
GB:AL009126 percent identity: 34.48; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AB98239.1"

## Alignment Scores:

Pred. No.: 0.114 Length: 10632  
Score: 112.00 Matches: 20  
Percent Similarity: 53.09% Conservative: 23  
Best Local Similarity: 24.69% Mismatches: 38  
Query Match: 21.79% Indels: 0  
DB: 1 Gaps: 0

US-10-089-514-4 (1-103) x U67480 (1-10632)

QY 1 MetThrGluGlnAsnGluLeuArgValArgValSerLeuValAspAlaLeuAspGlyThr 20  
Db 1932 GTTACATGATAGAGAACTCTCGAATTAAGAGAGAGTTATGATGACATTAAG 1873  
QY 21 LeuLeuAspThrValArgArgArgIleAspLeuGlyValArgIleAlaArgTyrIysSer 40  
Db 1872 ATATTAAAGCTAATTCCTTAAGAAAGAAATGTTAGCTAAGAGTACCTGATMAAAAT 1813  
QY 41 ArgHisGlyValProMetMetGlnProGlyArgValSerLeuValAspAlaGlyAla 60  
Db 1812 CAGCTTGATATTCCTTTAAGACGCCAGAAAGAAATAATATATACGATAGATTAAGA 1753  
QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIle 80  
Db 1752 AAACCTTGTAAGACATTAACGTTGATGAATAATTTGGCATTTAAATATTTCAAAATACTT 1693  
QY 81 Ile 81  
Db 1692 ATA 1690

RESULT 14  
AR271569\_12  
WPCOMMENT

Sequence split into 17 fragments LOCUS AR271569 Accession AR271569

Fragment Name	Begin	End
AR271569_00	1	110000
AR271569_01	100001	210000
AR271569_02	200001	310000
AR271569_03	300001	410000
AR271569_04	400001	510000
AR271569_05	500001	610000
AR271569_06	600001	710000
AR271569_07	700001	810000
AR271569_08	800001	910000
AR271569_09	900001	1010000
AR271569_10	1000001	1110000
AR271569_11	1100001	1210000
AR271569_12	1200001	1310000
AR271569_13	1300001	1410000
AR271569_14	1400001	1510000
AR271569_15	1500001	1610000

AR271569\_16 1600001 1664976  
Continuation (13 of 17) of AR271569 from base 1200001 (AR271569 Sequence 1 from patent L

## Alignment Scores:

Pred. No.: 1.36 Length: 110000  
Score: 112.00 Matches: 20  
Percent Similarity: 53.09% Conservative: 23  
Best Local Similarity: 24.69% Mismatches: 38  
Query Match: 21.79% Indels: 0  
DB: 6 Gaps: 0

US-10-089-514-4 (1-103) x AR271569\_12 (1-110000)

QY 1 MetThrGluGlnAsnGluLeuArgValArgValSerLeuValAspAlaLeuAspGlyThr 20  
Db 38358 GTTACATGATAGAGAACTCTCGAATTAAGAGAGAGTTATGATGACATTAAG 38417  
QY 21 LeuLeuAspThrValArgArgArgIleAspLeuGlyValArgIleAlaArgTyrIysSer 40  
Db 38418 ATATTAAAGCTAATTCCTTAAGAAAGAAATGTTAGCTAAGAGTACCTGATMAAAAT 38477  
QY 41 ArgHisGlyValProMetMetGlnProGlyArgValSerLeuValAspAlaGlyAla 60  
Db 38478 CAGCTTGATATTCCTTTAAGACGCCAGAAAGAAATAATATATACGATAGATTAAGA 38537  
QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIle 80  
Db 38538 AAACCTTGTAAGACATTAACGTTGATGAATAATTTGGCATTTAAATATTTCAAAATACTT 38597  
QY 81 Ile 81  
Db 38598 ATA 38600

## RESULT 15

AE000858/c 13242 bp DNA linear BCT 19-UTN-2002

LOCUS Methanobacterium thermoautotrophicum from bases 725908 to 739149

DEFINITION (section 64 of 148) of the complete genome.

ACCESSION AE000858 AE000666

VERSION AE000858.1 GI:2621885

## KEYWORDS

Methanobacteriaceae; Methanobacteriales; Methanobacterium

Methanobacterium thermoautotrophicus str. Delta H

Methanobacterium thermoautotrophicum str. deltaH

Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;

Methanobacteriaceae; Methanobacter.

1 (bases 1 to 13242)

Smith,D.R., Doucette-Stamm,L.A., Deloughery,C., Lee,H., Dubois,J.,

Aldredge,T., Bashirzadeh,R., Blakey,D., Cook,R., Gilbert,K.,

Harrison,D., Hoang,L., Keagle,P., Lum,W., Pothier,B., Qiu,D.,

Spadafora,R., Vicaire,R., Wang,Y., Wierzbowski,J., Gibson,R.,

Jiwani,N., Caruso,A., Bush,D. and Reeve,J.N.

Complete genome sequence of Methanobacterium thermoautotrophicum

deltaH: Functional analysis and comparative genomics

J. Bacteriol. 179 (22), 7135-7155 (1997)

## TITILE

## JOURNAL

## MEDLINE

## PUBMED

## AUTHORS

## TITLE

## JOURNAL

## FEATURES

## source

## Location/Qualifiers

## 1..13242

## /organism="Methanobacterium thermoautotrophicus str.

## Delta H"

## /mol\_type="genomic DNA"

## /strain="Delta H"

## /db\_xref="taxon:187420"

## /clone="WTH"

## /note="synonym: Methanobacterium thermoautotrophicum str.

## deltaH"



gene  
/gene="MTH796"  
62..676  
CDS  
/gene="MTH796"  
/note="Function Code:14.01 - Unknown, Conserved protein;  
similar to, pIR:LN:G64397 AC:G64397, p()=5.5E-17, pid=30%"  
/codon\_start=1  
/transl\_table=11  
/product="conserved protein"  
/protein\_id="AAB85296.1"  
/db\_xref="GI:2621886"  
/translation="MKIIDLTGKIEDSPVFPQDPYKLRIESSDADYITSSILGKH  
TGTIDAPLHAHSGALTVDGIGLEELTGEGLVSGGSIINGGDIADAVIRGTSMSSGSE  
GYFRDYGIKRLAEELVEHVGCVDPSPGSETDIHKLLKXGIVIVENTWT  
DLPKPKRLFVVPISVAEASPARVFAVTSSGSSGSSSDRSSFOR"  
complement (680..1216)  
gene  
/gene="MTH797"  
complement (680..1216)  
CDS  
/gene="MTH797"  
/note="Function Code:14.00 - Unknown, ; similar to,  
pIR:LN:S33068 AC:S33068, p()=0.055, pid=34%"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAB85297.1"  
/db\_xref="GI:2621887"  
/translation="MERDTMRFRMRDSREKQKULYKACMEISKLKRDJAKCRNSDN  
CERVKLELDALKTDRRIDELERTMAEKDVIQELKGIADKESRITDLRYEYFQA  
LTQPEKDLTSFQSIYRLLPDERATTEMLDFINEIGFDLKENNVQILRLMERKG  
YFRSVSEGRRTLMEKYKR"  
complement (1431..1991)  
gene  
/gene="MTH798"  
complement (1431..1991)  
CDS  
/gene="MTH798"  
/note="Function Code:12.09 - Cell Processes, Protein and  
peptide secretion ; similar to, sp:LN:Y374\_METUA  
AC:Q57819, p()=1.7E-13, pid=28%"  
/codon\_start=1  
/transl\_table=11  
/product="bactitracin transport permease related protein"  
/protein\_id="AAB85298.1"  
/db\_xref="GI:2621888"  
/translation="MHPTLLGPDLSVLYFPNSVLAAPFLDPLMLITFGGTOAFWVL  
CLILYLGSEDEREAPMALTAIVLGPFLSEALKMVIARPRPEVIGGMARNAVAGGY  
SMPSGAAVAFAGFISLYFLGRPFITIIASLVGISRTYLGHYSDVLAGAVLGL  
CAPTALKIDERVACFGLCRPDRLOQN"  
complement (1998..3041)  
gene  
/gene="MTH799"  
complement (1998..3041)  
CDS  
/gene="MTH799"  
/note="Function Code:5.01 - L-Amino Acid Metabolism,  
Alanine--aspartate and glutamate metabolism ; similar to,  
sp:LN:DHA5\_METUA AC:Q57658, p()=2.4E-96, pid=55%"  
/codon\_start=1  
/transl\_table=11  
/product="aspartate-semialdehyde dehydrogenase"  
/protein\_id="AAB85299.1"  
/db\_xref="GI:2621889"  
/translation="MVNVGLGATGAVGQRFIOMLDKHPFELTTLAASRSAGKPYG  
EVANWYLDCEPESVDMEVETDPSAVGVDILFSLPADVARKVEPKFAKYIAS  
NASAMMEBPVPLVPEVNPEPLDLEVOGRRGSGFITVNPNCSTIALTLTKPY  
DAYTKRVVSTQWASGAGNGVPSMAIINDNVPRFSGSEKILETTTLLLELDBG  
VYKPAFGVSGASCHRVAVVDGHTBAVFTLEDDFDDVDREANDKFRGLQGLHSA  
PEKPVVVRDEENRPQPMRDMDGMAVTVGRLEDAAFKNSLRVYLVGHNTVGAAG  
ASILNAELINEIL"  
complement (3054..3875)  
gene  
/gene="MTH800"  
complement (3054..3875)  
CDS  
/gene="MTH800"  
/note="Function Code:5.07 - L-Amino Acid Metabolism,  
Lysine biosynthesis ; similar to, pIR:LN:F64352 AC:F64352,  
p()=8E-80, pid=56%"  
/codon\_start=1

/transl\_table=11  
/product="dihydrodipicolinate reductase"  
/protein\_id="AAB85300.1"  
/db\_xref="GI:2621890"  
/translation="MIRVAVTGACGMSGCIIRRVLEEDEMLVAALTEARCTPLRGD  
IGBFTRGSVGVETDASNLADTAEPEVLDVFPVASAATITIKTSTAGVNLVG  
TTGSEEMQTVADCIERSGSRVAVIAPNMAVGFVGVADLAPILSDYDEIEMH  
HRHKDAPSGTAVRALVETISATGRASAEVAVHGRGGLTGERSDDELGHVAVGGDVI  
GDIVYFADGSELEIVHRAHSDATIGVYIRAIRTEDAEPGRINDMGVGLIN"  
complement (3928..4779)  
gene  
/gene="MTH801"  
complement (3928..4779)  
CDS  
/gene="MTH801"  
/note="Function Code:5.07 - L-Amino Acid Metabolism,  
lysine biosynthesis ; similar to, sp:LN:DAA9\_METUA  
AC:Q57695, p()=1.5E-69, pid=50%"  
/codon\_start=1  
/transl\_table=11  
/product="dihydrodipicolinate synthase"  
/protein\_id="AAB85301.1"  
/db\_xref="GI:2621891"  
/translation="MKIEGTVAAMVTPTFEDVDVADAGLREININLYIENGVDGLVAG  
TTGESATITHEEQRMIDIVDENGRVRTVAGAGSNSSREAGLVYARADAGDAL  
VTPYVYKQPPHGLIHHYTMLEBAADIPLIYVVPRTGIDIDVFAELAKLDGIG  
IKASPDLDKVSMLRSRLMDIGDDPTVLSGNLITLPMISMAEGVISVAVVDPAR  
MSRLVNEALSGDPESAMKTHYEIYSLMKVLIETSNVYPVKEALNMGKARGCEDAGC  
TPAGCKP"  
complement (4776..5996)  
gene  
/gene="MTH802"  
complement (4776..5996)  
CDS  
/gene="MTH802"  
/note="Function Code:5.01 - L-Amino Acid Metabolism,  
Alanine--aspartate and glutamate metabolism ; similar to,  
sp:GI:e255104:g1770047 LN:BS275208, p()=1.4E-36, pid=29%"  
/codon\_start=1  
/transl\_table=11  
/product="aspartokinase II alpha subunit"  
/protein\_id="AAB85302.1"  
/db\_xref="GI:2621892"  
/translation="MELIVAKFGGTSINGRRIKKARSVYKMKRKYVVVVSAIN  
KTTDELQIVDEAMDEVAITEKQAEIVSNGMESVIFSSAIALGVKSYIDPFME  
WPIITDSNLNAKVADEATEBEKRELLKLDGDIIPVCGFLGRDNGYTTTGROGS  
DITAFILGICAKDEVIIYTDVGWSTDPKIOGAKKDKISVEMRDLATHTGAOVL  
HRAALXKPPDIKAKIIGFBHGLSAPGEIIPCSKRYKTTTLPDISVAVVGS  
KIINKPGIARLTSRLAENSINITIGISTGNSVTIIVDKKDADEARHLHDVYIADD  
LSLSLGRDIAMTITISSPDIPTPGIISETIKPLRDNDAIVISSQTSVIVFDVN  
DGKAYELVAGVLE"  
complement (6046..6234)  
gene  
/gene="MTH803"  
complement (6046..6234)  
CDS  
/gene="MTH803"  
/note="Function Code:10.04 - Metabolism of Macromolecules,  
Ribosomal proteins; similar to, sp:LN:R17E\_METUA  
AC:P54026, p()=3.3E-12, pid=46%"  
/codon\_start=1  
/transl\_table=11  
/product="ribosomal protein S17"  
/protein\_id="AAB85303.1"  
/db\_xref="GI:2621893"  
/translation="MGNIRTSFVKRIAKEMIETHPEKFTDDPDTNKKLVBEFSTGK  
HLRNKAGIITRISQK"  
complement (6236..6532)  
gene  
/gene="MTH804"  
complement (6236..6532)  
CDS  
/gene="MTH804"  
/note="Function Code:5.11 - Tyrosine metabolism ; similar  
to, pIR:LN:G64330 AC:G64330, p()=8.9E-10, pid=32%"  
/codon\_start=1  
/transl\_table=11  
/product="chorismate mutase, subunit A"  
/protein\_id="AAB85304.1"  
/db\_xref="GI:2621894"  
/translation="MDEVARAEVLRSRQKIDIDRDIDLITSRIALAREIAKEV



gene  
LGMELIDPERELQIERTRKIARENGIDENKILTELMKILMDSKTEQKEMLRQ"  
complement (6573. .7442)  
/gene="MT805"  
complement (6573. .7442)  
/note="Function Code:14.01 - Unknown, Conserved protein;  
similar to, p1:LN:G64479 AC:G64479, p1=1.8E-36, pid=31%"  
/codon\_start=1  
/transl\_table=11  
/product="conserved protein (homoserine kinase related)"  
/protein\_id="AAB85305.1"  
/db\_xref="GI:2621895"  
/translation="MNVVKKTVRPGSATVYNAIARGSGAFGILRVEAEALIDSG  
VEICISREGADTSLMELCTRMVLEHYGVDAVRVTSDDLPAVSGLSSSAASNAVTMA  
VSSILSDERFLQPMEDFEMLMNAVDASLQAGVSVTGAYDDASAFYGLITVDNMERR  
ILIREPMENQKVLIVMPDRKSLTAQSDVPRMKLLAPWDMAPREVLDGRVHSALTNG

Alignment Scores:  
Pred. No.: 0.268 Length: 13242  
Score: 109.50 Matches: 26  
Percent Similarity: 54.55% Conservative: 22  
Best Local Similarity: 29.55% Mismatches: 39  
Query Match: 21.30% Indels: 1  
DB: 1 Gaps: 1

US-10-089-514-4 (1-103) x AE000858 (1-13242)

QY 7 LeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrValArg 26  
Db 6505 CTCAGAGATCCGACGACGATAGATGATGACGAGACATACCTGACCTCATTAACC 6446  
QY 27 ArgArgIleAspLeuGlyValArgIleAlaArgTyrLysSerArgHisGlyValProMet 46  
Db 6445 TCAAGGATAGCCCTTGACGCGGAGATGACGAGCCAGAGAGTCCCTGGATGAGATC 6386  
QY 47 MetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyrAlaAlaAspHis 66  
Db 6385 CTGGATCCTGAAAGGAACTCCAGATTATAGAGAGACCCGAAAATCGCCAGGAAAAT 6326  
QY 67 GlyLeuAspGlySerPheLeuValAsnLeuTyrAspValIleIleThrGlyMetCysArg 86  
Db 6325 GGATTCATGAGATAAATTACAGAACTGATGAGATTCTGATG--GATTCAGTAA 6269  
QY 87 ValGluAspLeuValMetSerArg 94  
Db 6268 ACTGAACAGAAAGAAATGTTAAG 6245

Search completed: October 5, 2005, 15:30:35  
Job time: 1687.4 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 4, 2005, 21:38:23 ; Search time 180.876 Seconds  
(without alignments)  
3371.006 Million cell updates/sec

Title: US-10-089-514-4

Perfect score: 514  
Sequence: 1 MFEQNELQRLRAELDALDGT.....MCRVEDLVMSRESLTAEDDR 103

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+2n.model -DEV=x1h  
-Q=/cgn2\_1/USPTO.spool\_h/US10089514/runat\_04102005\_105743\_8027/app.query.fasta\_1.1621  
-DB=N\_Geneseq -QMT=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOFC=0 -LOOPEXT=0  
-UNITS-bits -START=1 -END=-1 -MATRIX=blcsm62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NORM=ext -HEADING=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10089514.CCGN\_1.1.860.@runat\_04102005\_105743\_8027 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_16Dec04:\*  
1: geneseqn1960s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	514	100.0	312	4	AAf86410
2	514	100.0	312	8	AAf86410 4-amino-4
3	486.5	94.6	3305	8	AAf86410 1 S venezue
4	439	85.4	12391	8	ABZ69799 Plasmid P
5	256	49.8	387	2	AAf86410 1 S venezue

6	256	49.8	2888	2	AAT59268	AAT59268 Streptococ
7	120	23.3	360	10	ACF71772	ACF71772 Phototrab
8	120	23.3	110000	10	ACF67367_49	Continuation (50 o
9	120	23.3	110000	10	ACF65387_1	Continuation (2 of
10	112	21.8	110000	2	AAW21209_12	Continuation (13 o
11	109	21.2	94968	11	ADM27081_16	Continuation (17 o
12	106	20.6	1135	10	ADFO0853	AdFO0853 Bacteri
13	105	20.4	1305	11	ACF98708	ACF98708 Klebsiell
14	102	19.8	1122	8	AAf86410	AAf86410
15	102	19.8	1122	12	ADJ98178	ADJ98178 Pantocoe
16	102	19.8	1122	6	ABN83449	ABN83449 Erwinia h
17	101	19.6	1253	1	AAW71054	AAW71054 Optimized
18	101	19.6	2058	1	AAW71053	AAW71053 Escherich
19	101	19.6	3659	1	AAW71109	AAW71109 phea arof
20	101	19.6	4189	1	AAW71111	AAW71111 phea arof
21	96.5	18.8	1833	6	AAf40136	AAf40136 Isoprenol
22	96	18.7	110000	6	ABN66826	ABN66826 Streptoco
23	96	18.3	339	6	ABN71527_05	Continuation (6 of
24	94	18.3	339	6	ABN69596	ABN69596 Streptoco
25	93	18.1	1131	11	ACH98777	ACH98777 Klebsiell
26	92	17.9	1122	10	ACF69143	ACF69143 Phototrab
27	92	17.9	110000	10	ACF67367_20	Continuation (21 o
28	92	17.9	243072	10	ACF65382	ACF65382 Phototrab
29	90	17.5	649	10	ACF66455	ACF66455 Phototrab
30	90	17.5	1122	6	ABN83448	ABN83448 Escherich
31	90	17.5	1122	8	AAf50753	AAf50753 Escherich
32	90	17.5	1158	10	ACF69141	ACF69141 Phototrab
33	90	17.5	1238	6	ABA99612	ABA99612 E. coli t
34	90	17.5	1238	6	AAf46032	AAf46032 E. coli t
35	90	17.5	110000	10	ACF67367_20	Continuation (21 o
36	90	17.5	243072	10	ACF65382	ACF65382 Phototrab
37	89.5	17.4	1104	8	ACA27476	ACA27476 Prokaryot
38	89.5	17.4	18096	4	AAf59588	AAf59588 Propionib
39	89.5	17.4	18096	8	ACF64517	ACF64517 Phototrab
40	89	17.3	652	3	AAf74584	AAf74584 Human ORF
41	89	17.3	652	6	ABN18494	ABN18494 Human ORF
42	88.5	17.2	804	11	ABD06210	ABD06210 Pseudomon
43	88.5	17.2	960	11	ABD06101	ABD06101 Pseudomon
44	88.5	17.2	1653	11	ABD06120	ABD06120 Pseudomon
45	88.5	17.2	1791	11	ABD06080	ABD06080 Pseudomon

## ALIGNMENTS

RESULT 1	
ID	AAf86410 standard; DNA; 312 BP.
XX	
AC	AAf86410;
XX	
DT	22-JUN-2001 (first entry)
XX	
DE	4-amino-4-deoxychorismic acid mutase coding sequence.
XX	
KM	Metabolite; benzene; chorismic acid; p-aminophenylpyruvic acid;
KW	4-amino-4-deoxychorismic acid mutase; enzyme; pabB; ds.
XX	
OS	Streptomyces venezuelae.
XX	
FT	Key
FT	Location/Qualifiers
FT	CDS
FT	1..312
FT	/tag= a
FT	/product= "4-amino-4-deoxychorismic acid mutase"
XX	
PD	WO200123542-A1.
XX	
PD	05-APR-2001.
XX	
PF	29-SEP-2000; 2000MO-DP006783.
XX	
PR	29-SEP-1999; 99JP-00276314.
XX	
PA	(MEIJ ) MEIJI SEIKA KAISHA LTD.



XX Yanai K, Okakura K, Yaeuda S, Watanabe M, Miyamoto K, Midoh N;  
PI Murakami T;  
XX  
DR WPI: 2001-290517/30.  
DR P-PSDB; AAB82072.  
XX  
PT Transformant producing secondary metabolite modified with functional  
PT group e.g. benzene with nitrogen-containing substituent at para-position,  
PT PF1022, with ease at low cost, for application in pharmaceuticals and  
PT agrochemicals.  
XX  
PS Claim 13; Page 65; 83pp; Japanese.  
XX  
CC The present invention relates to a transformant having been modified so  
CC as to produce a secondary metabolite. The secondary metabolite has a  
CC benzene ring skeleton free from substitution at the para-position by a  
CC nitrogen-containing functional group, thereby enabling the production of  
CC a secondary metabolite with a benzene ring skeleton substituted at the  
CC para-position by a nitrogen-containing group. The transformant organism  
CC of the present invention has been produced by transferring a gene  
CC participating in the biosynthesis pathway from chorismic acid into p-  
CC aminophenylpyruvic acid. The present sequence is the coding sequence for  
CC 4-amino-4-deoxychorismic acid mutase (pabp), from Streptomyces  
CC venezuelae. pabp participates in the biosynthesis pathway from chorismic  
CC acid into p-aminophenylpyruvic acid, and so the pabp gene can be used to  
CC produce the transformant of the present invention. The transformant can  
CC be used to produce metabolites for application in pharmaceuticals,  
CC veterinary drugs and agrochemicals  
XX  
SQ Sequence 312 BP; 54 A; 106 C; 111 G; 41 T; 0 U; 0 Other;  
XX  
Alignment Scores:  
Pred. No.: 2,38e-61 Length: 312  
Score: 514.00 Matches: 103  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0  
US-10-089-514-4 (1-103) x AAF86410 (1-312)  
QY 1 MetThGlGlnGlnGluGluGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20  
DB 1 ATGACCGAGCAGAACGAGCTGACGCGGCTGCGCGGAGCTCGACCCCTCGACGGGAGC 60  
QY 21 LeuLeuAspThrValArgArgArgAlaAspLeuGlyValArgAlaArgTyrIysSer 40  
DB 61 CTCCTGGACACGGTGGCGGCGCCGATCGACTCGGTGTCGCGATCGCGGATCAAGTCC 120  
QY 41 ArgHisGlyValProMetMetGlnProGlyArgValSerLeuValIysAspArgAlaAla 60  
DB 121 CGGCAAGGGGTCCGATGATGACGCCCGCGGCTGAGCTGTCAAGACAGCGGCCGCC 180  
QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGlySerPheLeuValAsnLeuTyrAspValIle 80  
DB 181 CGCTACGGCGCGCAGCACCGGCTCGACGAATCGTCTCGGTGAACCTTACGACGTGATC 240  
QY 81 IleThGlnMetCysArgValGluAspLeuValMetSerArgGluSerLeuThrAlaGlu 100  
DB 241 ATCAAGAAATGTGGCGGTGAGGACCTGTGATGAGCGGAGAGCCTGACGGCGCAG 300  
QY 101 AspArgArg 103  
DB 301 GACCGGCGG 309  
RESULT 2  
AAL50181  
ID AAL50181 standard; DNA; 312 BP.  
XX  
AC AAL50181;  
XX  
DT 28-JAN-2003 (first entry)

XX  
DE S venezuelae PF1022 substance gene #2.  
XX  
XX Substance PF1022; biosynthesis; transgenic; chorismic acid; anthelmintic;  
KW phenylalanine; p-aminophenylpyruvic acid; pharmaceutical;  
KW veterinary drug; gene; ds.  
XX  
XX Streptomyces venezuelae.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..312  
FT /tag=a  
FT /product= "PF1022 substance"  
XX  
XX MO200277244-A1.  
XX  
XX PD 03-OCT-2002.  
XX  
XX PF 22-MAR-2002; 2002MO-JP002782.  
XX  
XX PR 22-MAR-2001; 2001JP-00082227.  
XX  
XX PA (MEIJI SEIKA KAISHA LTD.  
XX PI Yanai K, Sumida N, Watanabe M, Moriya T, Murakami T;  
XX DR WPI: 2003-018934/01.  
XX DR P-PSDB; AAO19564.  
XX  
XX PT Novel biosynthesis gene-transferred transformants for producing PF1022  
XX PT substance derivatives by fermentation, as pharmaceuticals or veterinary  
XX PT drugs with anthelmintic activity.  
XX  
XX PS Claim 15; Page 85; 116pp; Japanese.  
XX  
XX CC The present invention relates to transformants capable of producing  
XX CC PF1022 substance derivatives. These were obtained by transferring a gene  
XX CC participating in the biosynthesis pathway from chorismic acid to p-  
XX CC aminophenylpyruvic acid (biosynthesis gene) into a phenylalanine-  
XX CC requiring host derived from an organism producing the PF1022 substance.  
XX CC The transformants are producing PF1022 substance derivatives by  
XX CC fermentation, for use as pharmaceuticals or veterinary drugs. The present  
XX CC sequence is a substance PF1022 coding sequence from Streptomyces  
XX CC venezuelae  
SQ Sequence 312 BP; 54 A; 106 C; 111 G; 41 T; 0 U; 0 Other;  
XX  
Alignment Scores:  
Pred. No.: 2,38e-61 Length: 312  
Score: 514.00 Matches: 103  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0  
US-10-089-514-4 (1-103) x AAL50181 (1-312)  
QY 1 MetThGlGlnGlnGluGluGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20  
DB 1 ATGACCGAGCAGAACGAGCTGACGCGGCTGCGCGGAGCTCGACCCCTCGACGGGAGC 60  
QY 21 LeuLeuAspThrValArgArgArgAlaAspLeuGlyValArgAlaArgTyrIysSer 40  
DB 61 CTCCTGGACACGGTGGCGGCGCCGATCGACTCGGTGTCGCGATCGCGGATCAAGTCC 120  
QY 41 ArgHisGlyValProMetMetGlnProGlyArgValSerLeuValIysAspArgAlaAla 60  
DB 121 CGGCAAGGGGTCCGATGATGACGCCCGCGGCTGAGCTGTCAAGACAGCGGCCGCC 180  
QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGlySerPheLeuValAsnLeuTyrAspValIle 80  
DB 181 CGCTACGGCGCGCAGCACCGGCTCGACGAATCGTCTCGGTGAACCTTACGACGTGATC 240



Qy 81 11ethrGlumecCysaRgVa1g1uAspLeuValmetSerArg1uSerLeuThrAlaGlu 100  
 |||  
 Db 241 ATACGAGATGTCTCCGCTGAGAGACTGTGATGACCCGAGAGAGCTTACGAGCCAG 300  
 Qy 101 AspaRgArg 103  
 |||  
 Db 301 GACCGCGCG 309

RESULT 3  
 ABZ69799  
 ID ABZ69799 standard; DNA; 3305 BP.  
 XX  
 AC ABZ69799;  
 XX  
 DT 08-APR-2003 (first entry)  
 XX  
 DE Plasmid papABC.  
 XX  
 KM DHFR; translation; orthogonal tRNA; O-tRNA; O-RS; CAT; TyrRS;  
 KM orthogonal aminoacyl tRNA synthetase; unnatural amino acid;  
 KM chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;  
 KM biosynthesis; p-aminophenylalanine; pAF; choriastmate.  
 XX  
 OS Synthetic.  
 OS  
 PN W0200285923-A2.  
 XX  
 PD 31-OCT-2002.  
 XX  
 PF 19-APR-2002; 2002WO-US012465.  
 XX  
 PR 19-APR-2001; 2001US-0285030P.  
 PR 06-FEB-2002; 2002US-0355514P.  
 XX  
 PA (SCRI ) SCRIPPS RES INST.  
 XX  
 PI Schultz P, Wang L, Anderson JC, Chin JWK, Liu DR, Magliery TJ;  
 PI Meggers EL, Wehl RA, Pasternak M, Santoro SW, Zhang Z;  
 DR WPI; 2003-120430/11.  
 XX  
 XX Composition useful for producing protein comprising unnatural amino acid,  
 PT has translation system comprising orthogonal tRNA and orthogonal  
 PT aminoacyl tRNA synthetase.  
 PS  
 PS Example 4; Page 127-128; 188bp; English.  
 XX  
 CC The invention relates to a novel composition comprising a translation  
 CC system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl  
 CC tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O  
 CC -tRNA with at least one unnatural amino acid in the translation system  
 CC and the O-tRNA recognises at least one selector codon. A composition of  
 CC the invention is useful for producing at least one protein comprising at  
 CC least one unnatural amino acid. The protein is the Asp127AG mutant of  
 CC chloramphenicol acetyltransferase (CAT), the Tyr163TAG mutant of mouse  
 CC dihydrofolate reductase (DHFR), or a Tyr163TAG mutant of mouse  
 CC dihydrofolate reductase comprising CHOOH His6tag. The unnatural amino  
 CC acid is provided exogenously. The translation system is a cell and the  
 CC unnatural amino acid is biosynthesised by the cell. The present sequence  
 CC represents a plasmid containing the individual genes papABC that encode  
 CC the enzymes used to carry out the conversion of choriastmate to the  
 CC unnatural amino acid p-aminophenylalanine (pAF)  
 XX  
 SQ Sequence 3305 BP; 445 A; 1327 C; 1121 G; 412 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 3,696-56 Length: 3305  
 Score: 486.50 Matches: 103  
 Percent Similarity: 98.10% Conservative: 0  
 Best Local Similarity: 98.10% Mismatches: 2  
 Query Match: 94.65% Indels: 0  
 DB: 8 Gaps: 1

US-10-089-514-4 (1-103) x ABZ69799 (1-3305)  
 Qy 1 MetThrGluGluAsnGluLeuGln---ArgLeuAraGlaGluLeuAspAlaLeuAspG1 19  
 |||  
 Db 2990 ATGACCCAGCAGACGAGAGCTGACAGTTCGCGCTGCGCGAGCTCGACCCCTCGACGG 3049  
 Qy 19 YThrLeuLeuAspThrValAraArgArg11eAspLeuG1VAlArg11eAlaArgTyrLy 39  
 |||  
 Db 3050 GACGCTTCTGACACGGGTGGCGCGCGCATGACCTCGGTGTCCGACATCCGCGGTACAA 3109  
 Qy 39 sSerArgHisG1VAlPrometMetGlnProG1ArgValSerLeuVal1yAspArgAl 59  
 |||  
 Db 3110 GTCCCGGACAGCGGCTCCCGATGATGACGCCCGCGCGGTCAAGTGTCAAGACAGCGC 3169  
 Qy 59 aAlaArgTyrAlaAlaAspHisG1LeuAspG1uSerPheLeuValAsnLeuTyrAspVa 79  
 |||  
 Db 3170 CGCCCGCTACCGCGCCGACACCGCTCGACGATCTTCCTGTAACCTCTACGACGT 3229  
 Qy 79 11e11ethrGlumecCysaRgVa1g1uAspLeuValmetSerArg1uSerLeuThrAl 99  
 |||  
 Db 3230 GATCATCAGCAGAGATGTGCGCGTGAAGACTGTGTGATGACCGGAGAGCCTGACGGC 3289  
 Qy 99 aGluAspArgArg 103  
 |||  
 Db 3290 CGAGGACCGCGCG 3302

RESULT 4  
 ABZ69798  
 ID ABZ69798 standard; DNA; 12391 BP.  
 XX  
 AC ABZ69798;  
 XX  
 DT 08-APR-2003 (first entry)  
 XX  
 DE Plasmid plasc-papabc.  
 XX  
 KM DHFR; translation; orthogonal tRNA; O-tRNA; O-RS; CAT; TyrRS;  
 KM orthogonal aminoacyl tRNA synthetase; unnatural amino acid;  
 KM chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;  
 KM biosynthesis; p-aminophenylalanine; pAF.  
 XX  
 OS Synthetic.  
 OS  
 PN W0200285923-A2.  
 XX  
 PD 31-OCT-2002.  
 XX  
 PF 19-APR-2002; 2002WO-US012465.  
 XX  
 PR 19-APR-2001; 2001US-0285030P.  
 PR 06-FEB-2002; 2002US-0355514P.  
 XX  
 PA (SCRI ) SCRIPPS RES INST.  
 XX  
 PI Schultz P, Wang L, Anderson JC, Chin JWK, Liu DR, Magliery TJ;  
 PI Meggers EL, Wehl RA, Pasternak M, Santoro SW, Zhang Z;  
 DR WPI; 2003-120430/11.  
 XX  
 XX Composition useful for producing protein comprising unnatural amino acid,  
 PT has translation system comprising orthogonal tRNA and orthogonal  
 PT aminoacyl tRNA synthetase.  
 PS  
 PS Example 4; Page 124-127; 188bp; English.  
 XX  
 CC The invention relates to a novel composition comprising a translation  
 CC system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl  
 CC tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O  
 CC -tRNA with at least one unnatural amino acid in the translation system  
 CC and the O-tRNA recognises at least one selector codon. A composition of  
 CC the invention is useful for producing at least one protein comprising at  
 CC least one unnatural amino acid. The protein is the Asp127AG mutant of  
 CC chloramphenicol acetyltransferase (CAT), the Tyr163TAG mutant of mouse



CC dihydrofolate reductase (DHFR), or a Tyr163TNG mutant of mouse  
CC dihydrofolate reductase comprising CHOOH His6tag. The unnatural amino  
CC acid is provided exogenously. The translation system is a cell and the  
CC unnatural amino acid is biosynthesised by the cell. The present sequence  
CC represents a plasmid for use in the biosynthesis of p-aminophenylalanine  
CC (pAp) in vivo

XX SQ Sequence 12391 BP, 2830 A, 3588 C, 3346 G, 2627 T, 0 U, 0 Other;

## Alignment Scores:

Pred. No.:	7,79e-49	Length:	12391
Score:	439.00	Matches:	96
Percent Similarity:	94.12%	Conservative:	0
Best Local Similarity:	94.12%	Mismatches:	3
Query Match:	85.41%	Indels:	3
DB:	8	Gaps:	2

US-10-089-514-4 (1-103) x ABZ69798 (1-12391)

QY 1 MetThrGluGlnAsnGluLeuGln---ArgLeuArgAlaGluLeuAspAlaLeuAspG1 19  
DB 2055 ATGACCGAGCAGACGAGCTGCGGTGGCGCTGCCGCGAGACTCGACGCCCTCGACGG 2114  
QY 19 YhrLeuLeuAspThrValArgArgArgTleAspLeuGlyValArgTleAlaArgTyr1 39  
DB 2115 GACGCTTCGGACACGGTGGCGCGCATCGACCTCGGTGTCGATCGCGGGTACAA 2174  
QY 39 sSerArgHisGlyValProMetMetGlnProGlyArgValSerLeuValLysAspArg1 59  
DB 2175 GTCCCGGACGCGGTGCCATGATGCAAGCCCGCGGTGACCTGTCAAGACAGGGC 2234  
QY 59 aAlaArgTyrAlaAlaAspHisGlyLeuAspGlnSerPheLeuValAsnLeuTyrAspVa 79  
DB 2235 CGCCCGCTACGCGCGCGACACCGCGCTCGCAATGCTTCGTGTAACCTTACGACGT 2294  
QY 79 lIleIleThrGlnMetCysArgValAlaGluAspLeuValMetSerArgGluSerLeuThr1 99  
DB 2295 GATCATCAGGAGATGTGCCGCGTGAAGACCTGTGTATGAGC---CCGTATGTACTAA 2351  
QY 99 aglu 100  
DB 2352 GGAG 2355

RESULT 5  
AAT58552 standard; cDNA; 387 BP.

XX AC AAT58552;

DT 02-APR-1997 (first entry)

XX DE Streptomyces pristinaespiralis papB gene.

XX KM Streptogramin B; antibiotic; biosynthesis; pristinaamycin; virginiamycin;  
XX DMPAPA precursor; 4-dimethylamino-L-phenylalanine; papA; papM; papB;  
XX KM papC; isomerisation; aromatisation; N-methyltransferase; ds.

XX OS Streptomyces pristinaespiralis.

XX FH Key Location/Qualifiers  
FT mat\_peptide 1..387  
FT /\*tag= a  
FT /product= "PapB"

PN MO9601901-A1.

PD 25-JAN-1996.

PF 04-JUL-1995; 95MO-FR000889.

XX PR 08-JUL-1994; 94FR-00008478.

XX PA (RHON ) RHONE POUJENC RORER SA.

XX XX Blanc V, Thibaut D, Bamas-Jacques N, Blanche F, Crouzet J;  
PI Barriere J, Debussche L, Famechon A, Paris J, Dutruc-Rosset G;  
XX WPI: 1996-097631/10.  
DR P-PSDB; AAM11583.

PT New streptogramin B derivs. useful as antibiotics - produced by new  
PT mutants of Streptomyces having altered genes for streptogramin B  
PT biosynthesis.

PS Claim 18; Page 106-107; 146pp; French.

XX The papA and papM genes of S.pristinaespiralis are involved in the  
XX CC biosynthesis of 4-dimethylamino-L-phenylalanine (DMPAPA), a precursor for  
XX CC pristinaamycin IA. The region between these two genes was sequenced and  
XX CC two open reading frames were identified. The first (papC) was on the  
XX CC complementary strand and decodes to an amino acid sequence with homology  
XX CC to the region of E.coli TyrA which has been implicated in aromatisation  
XX CC reactions. The papC gene product is likely to be involved in a similar  
XX CC aromatisation of 4-deoxy 4-amino prephenate to give 4-amino  
XX CC phenylpyruvate during DMPAPA synthesis. The second open reading frame  
XX CC (papB) could be decoded to give a product with homology to the region of  
XX CC TyrA which has chorismate mutase activity. The papB gene product is  
XX CC likely to catalyse the isomerisation of 4-deoxy 4-amino chorismate to 4-  
XX CC deoxy 4-aminoprephenate acid. Disruption of the papB and papC genes can  
XX CC be used to produce strains of S.pristinaespiralis which are unable to  
XX CC produce the antibiotic pristinaamycin I but which may be able to produce  
XX CC new, modified forms of it

SQ Sequence 387 BP; 50 A; 187 C; 110 G; 40 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	1.03e-25	Length:	387
Score:	256.00	Matches:	51
Percent Similarity:	71.08%	Conservative:	8
Best Local Similarity:	61.45%	Mismatches:	24
Query Match:	49.81%	Indels:	0
DB:	2	Gaps:	0

US-10-089-514-4 (1-103) x AAT58552 (1-387)

QY 7 LeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrValArg 26  
DB 64 CTGACGCGCTGCGCGCGCGCTCGACGCGCGGACGCGCCCTGTGAGCCCGCCGCGC 123  
QY 27 ArgArgIleAspLeuGlyValArgTleAlaArgTyrLysSerArgHisGlyValProMet 46  
DB 124 ACAAGCTGAGATGCTGCTGCGCATCGCGGAGTACAGCGCGCTCCACGAGTCCGATG 183  
QY 47 MetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyrAlaAlaAspHis 66  
DB 184 ATGAGAGCCCGACCGATGCGCCAGTCCAGTCCAGCGCAACGCCCGCTACGCCCGCACGC 243  
QY 67 GlyLeuAspGlnSerPheLeuValAsnLeuTyrAspValIleIleThrGlnMetCysArg 86  
DB 244 GGCATCAGACCCCGCTTCTCGGCAACCTGTACGACACGATCATCCAGACACTGCCGC 303  
QY 87 ValGluAsp 89  
DB 304 CTCGAGGAC 312

RESULT 6  
AAT59268 standard; cDNA; 2888 BP.

XX AC AAT59268;

DT 02-APR-1997 (first entry)

XX DE Streptomyces pristinaespiralis papA and papM intergenic region.

XX KM Streptogramin B; antibiotic; biosynthesis; pristinaamycin; virginiamycin;



KM DMPAPA precursor; 4-dimethylamino-L-phenylalanine; papA; papM; papB;  
 KM papC; isomerisation; aromatisation; N-methyltransferase; de.  
 XX  
 OS Streptomyces pristinaeaepristalis.  
 FH  
 FH Key Location/Qualifiers  
 FT CDS 1..687  
 FT /\*tag= a  
 FT /product= "PapA"  
 FT /note= "C-terminal coding region only, i.e. a partial  
 open reading frame"  
 FT complement(949..1836)  
 FT CDS  
 FT /\*tag= b  
 FT /product= "PapC"  
 FT 1873..2262  
 FT /\*tag= c  
 FT /product= "PapB"  
 FT 2259..2888  
 FT /\*tag= d  
 FT /product= "PapM"  
 FT /note= "N-terminal coding region only, i.e. a partial  
 open reading frame"  
 FT  
 XX  
 XX MO9601901-A1.  
 XX  
 XX 25-JAN-1996.  
 XX  
 XX 04-JUL-1995; 95WO-FR000889.  
 XX  
 XX 08-JUL-1994; 94FR-00008478.  
 XX  
 XX (RHON ) RHONE POULENC RORER SA.  
 XX  
 XX Blanc V, Thibaut D, Bamas-Jacques N, Blanche F, Crouzet J;  
 PI Barriere J, Debussche L, Famechon A, Paris J, Dutruc-Rosset G;  
 XX  
 XX WPI; 1996-097631/10.  
 DR P-PSDB; AAM11582, AAM11583.  
 DR  
 PT New streptogramin B derive. useful as antibiotics - produced by new  
 PT mutants of Streptomyces having altered genes for streptogramin B  
 PT biosynthesis.  
 XX  
 XX Example 1; Page 102-104; 146pp; French.  
 XX  
 CC The papA and papM genes of *S. pristinaeaepristalis* are involved in the  
 CC biosynthesis of 4-dimethylamino-L-phenylalanine (DMPAPA), a precursor for  
 CC pristinamycin 1A. The region between these two genes was sequenced and  
 CC two open reading frames were identified. The first (papC) was on the  
 CC complementary strand and decodes to an amino acid sequence with homology  
 CC to the region of *E. coli* tyra which has been implicated in aromatisation  
 CC reactions. The papC gene product is likely to be involved in a similar  
 CC aromatisation of 4-deoxy 4-amino prephenate to give 4-amino  
 CC phenylpyruvate during DMPAPA synthesis. The second open reading frame  
 CC (papB) could be decoded to give a product with homology to the region of  
 CC tyra which has chorismate mutase activity. The papB gene product is  
 CC likely to catalyse the isomerisation of 4-deoxy 4-amino chorismate to 4-  
 CC deoxy 4-amino prephenate. Disruption of the papB and papC genes can be  
 CC used to produce strains of *S. pristinaeaepristalis* which are unable to  
 CC produce the antibiotic pristinamycin I but which may be able to produce  
 CC new, modified forms of it  
 XX  
 SQ Sequence 2888 BP; 390 A; 1319 C; 897 G; 282 T; 0 U; 0 Other;

QY 7 leuGlnArgleuArgAlaIleuAspAlaLeuAspGlyThrleuLeuAspTrpValArg 26  
 DB 1936 CTCGACGCGCTGGCCGCTCCCTCGACGCGCGGAGCGCCGCTCGACGCGCTCCG 1995  
 QY 27 ArgArgIleAspLeuGlyValArgIleAlaArgTyrlsSerArgIleGlyValPromet 46  
 DB 1996 ACAGCGCTGGACATCTGGCGCATCGGAGATACAGAGCGCTCCACGAGTCCGATG 2055  
 QY 47 MetGlnProGlyArgValSerleuValIlyAspArgAlaAlaArgTyrlaAlaAspHis 66  
 DB 2056 ATCCAGCCCGCCAGATGCGCCAGTCCAGCCAGCCAGCCGCTACAGCGCGGACAC 2115  
 QY 67 GlyLeuAspGlnSerPheLeuValAsnLeuTyrrAspValIleIleThrgIleMetCysArg 85  
 DB 2116 GGCATCGACCCCGCTTCGCGCACCTGTACACACATCATCCGAGACTGCGCG 2175  
 QY 87 ValGluAsp 89  
 DB 2176 CTCGAGGAC 2184  
 RESULT 7  
 ACF71772  
 ID ACF71772 standard; DNA; 360 BP.  
 XX  
 AC ACF71772;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Photorhabdus luminescens nucleotide sequence #10239.  
 XX  
 KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KW detection; food; gene expression; plant; animal; microorganism; toxin;  
 KW antibiotic; biopesticide; virulence factor; disease model; plague;  
 KW whooping cough; gene; ds.  
 XX  
 OS Photorhabdus luminescens.  
 XX  
 XX W0200294867-A2.  
 PN  
 XX 28-NOV-2002.  
 PD  
 XX  
 PF 07-FEB-2002; 2002MO-IB003040.  
 XX  
 PR 07-FEB-2001; 2001FR-00001659.  
 PR  
 XX (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Duchaud E, Taourit S, Glaeer P, Frangeul L, Kunet F, Danchin A;  
 PI Buchrieser C;  
 XX  
 DR WPI; 2003-148459/14.  
 DR  
 PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 PT  
 PS Claim 2; SEQ ID NO 10239; 1205bp; French.  
 PS  
 CC The invention relates to the isolation of genes and their encoded  
 CC proteins from Photorhabdus luminescens. The isolated sequences are  
 CC sources of probes and primers for detecting the genome of *P. luminescens*  
 CC and related species; to study polymorphisms; for gene analysis and for  
 CC detection/amplification of the genes. Antibodies (Ab) raised against the  
 CC polypeptides encoded by the genes are used for detection/identification  
 CC of *P. luminescens*, e.g. in foods. The genes, proteins, Ab and cells that  
 CC carry a gene-containing vector are used to select compounds that  
 CC modulate, regulate, induce or inhibit expression of the genes in plants,  
 CC animals or microorganisms other than *P. luminescens* and are able to alter  
 CC response or sensitivity to toxins and antibiotics produced by *P.*  
 CC luminescens. Cells transformed to express the genes are useful for  
 CC recombinant production of the proteins, particularly toxins and  
 CC antibacterials useful as insecticides, bactericides and fungicides. The



CC genes/proteins, vectors containing the genes and Ab are also useful  
CC therapeutically (to treat microbial infection by bacteria or fungi) that  
CC are sensitive to P. luumniscens-encoded toxins or antibiotics) and as  
CC biopesticides. Other uses of the genes and the proteins are as virulence  
CC factors and for identifying targets of human diseases for which P.  
CC luumniscens is a model (particularly plague and whooping cough). This  
CC sequence represents one of the isolated P. luumniscens genes

### Alignment Scores:

Pred. No.:	4.9e-07	Length:	360
Score:	120.00	Matches:	25
Percent Similarity:	55.06%	Conservative:	24
Best local Similarity:	28.09%	Mismatches:	40
Query Match:	23.35%	Indels:	0
DB:	10	Gaps:	0

US-10-089-514-4 (1-103) X ACF71772 (1-360)

Oy		MetThcIcludlnsnegluleuglnrArgleAaglaIagluleAspAlaleuasgcyLThr	20
Dd	34	ATATAATTAACAGGAAATATTACAACCTTTCACGGGAATCTTGAGATTATTAATATCAG	93
Oy	21	LeuleuAspThrValArgrArgrglleAspleuglyValArgiliAlarTyLysSer	40
Dd	94	ATTTGGCACTGCCTTGTGACGTATGAAGAATTTGTATGAANAATCGCAGGTAAACGG	153
Oy	41	ArgHlsBglyAlPrometmetcInPrOgiylArGlvalserLeuvalLvAspArGlAlala	60
Dd	154	GCGCAGAGATVTTCCGATGATGCCACCACAACGTATTACTCGTATTAAGAATATGTTCCGG	213
Oy	61	ArgTyrlalAlasphIsiglyLeuAspgIusePheleuValAsmLeuTyAspAlile	80
Dd	214	GATTAATTCACCGATTTTGGGTTACGGCGCTGAATTAATCTGAATCAATATTTCACGTGTT	273
Oy	81	IleThrGliumecysargValGIusp	89
Dd	274	ATTGAAGACCTTGGCGTCGAGAAGAA	300

RESULT 8			
Constitution (50 of 57) of ACF67367 from base 4900001 (Phototaxibius luminescens nucleotid			
WP	Fragment Name	Begin	End
WP	ACF67367_00	1	110000
WP	ACF67367_01	100001	210000
WP	ACF67367_02	200001	310000
WP	ACF67367_03	300001	410000
WP	ACF67367_04	400001	510000
WP	ACF67367_05	500001	610000
WP	ACF67367_06	600001	710000
WP	ACF67367_07	700001	810000
WP	ACF67367_08	800001	910000
WP	ACF67367_09	900001	1010000
WP	ACF67367_10	10000001	1110000
WP	ACF67367_11	1100001	1210000
WP	ACF67367_12	1200001	1310000
WP	ACF67367_13	1300001	1410000
WP	ACF67367_14	1400001	1510000
WP	ACF67367_15	1500001	1610000
WP	ACF67367_16	16000001	1710000
WP	ACF67367_17	17000001	1810000
WP	ACF67367_18	1800001	1910000
WP	ACF67367_19	1900001	2010000
WP	ACF67367_20	2000001	2110000
WP	ACF67367_21	2100001	2210000
WP	ACF67367_22	2200001	2310000
WP	ACF67367_23	2300001	2410000
WP	ACF67367_24	2400001	2510000
WP	ACF67367_25	2500001	2610000
WP	ACF67367_26	2600001	2710000
WP	ACF67367_27	27000001	2810000

WP	ACF67367_28	2800001	2910000
WP	ACF67367_29	2900001	3010000
WP	ACF67367_30	3000001	3110000
WP	ACF67367_31	3100001	3210000
WP	ACF67367_32	3200001	3310000
WP	ACF67367_33	3300001	3410000
WP	ACF67367_34	3400001	3510000
WP	ACF67367_35	3500001	3610000
WP	ACF67367_36	3600001	3710000
WP	ACF67367_37	3700001	3810000
WP	ACF67367_38	3800001	3910000
WP	ACF67367_39	3900001	4010000
WP	ACF67367_40	4000001	4110000
WP	ACF67367_41	4100001	4210000
WP	ACF67367_42	4200001	4310000
WP	ACF67367_43	4300001	4410000
WP	ACF67367_44	4400001	4510000
WP	ACF67367_45	4500001	4610000
WP	ACF67367_46	4600001	4710000
WP	ACF67367_47	4700001	4810000
WP	ACF67367_48	4800001	4910000
WP	ACF67367_49	4900001	5010000
WP	ACF67367_50	5000001	5110000
WP	ACF67367_51	5100001	5210000
WP	ACF67367_52	5200001	5310000
WP	ACF67367_53	5300001	5410000
WP	ACF67367_54	5400001	5510000
WP	ACF67367_55	5500001	5610000
WP	ACF67367_56	5600001	5618894

**Alignment Scores:**

Pred. No.:	0.00125	Length:	1100000
Score:	120.00	Matches:	25
Percent Similarity:	55.06%	Conservative:	24
Best Local Similarity:	28.09%	Mismatches:	40
Query Match:	23.35%	Indels:	0
DB:	10	Gaps:	0

US-10-089-514-4 (1-103) X ACF67367\_49 (1-110000)

OY	1	MechrIugInsgInsgIueugInsgIueAsgIaIglIueAspIaIeuhsgIythr	20
		::: ::::	::: ::::
Db	74070	ATAATTAACAAGGAATATTACAACCTTACAGGAATCTTTGGATGTTATTATATGCG	74129
OY	21	LeuIeuaSpThrValAlrGa-rGa-rglIeaSpIueugIValArgIleaIaArgTyIySer	40
		:::: ::::	:::: ::::
Db	74130	ATTTTGGCACTGCTTTCTGTAACGTATGAAGAAGTTGTGATGAATAATCGCTAGCTAAAGCG	74189
OY	41	ArgIhsIeIyValIProMetGtsInPrGgIYArgValSerLeuValIeAspIaIaIa	60
		::: ::::	::: ::::
Db	74190	GGCGAGATATTCGATGATGACGCCAACAGTATTAACCTCGTATTGAAATGTTCGCG	74249
OY	61	ArgTyIleaIaAspIhsIeugIleuAspIuSerPheIueValaIaIeuhIyAspValIle	80
		::: ::::	::: ::::
Db	74250	GATAAATCAACCGATTTTGGGTTACGCGCTGAATATATCGAATCAATATTCAACGCTT	74309
OY	81	IleThrgIueTcySarIyValGIueAp	89
Db	74310	ATTGCAAGACCTTGCCCTCGACGAAGA	74336

RESULT 9			
ACF65387_1			
Continuation (2 of 7) of ACF65387 from base 100001 (Photorhabdus luminescens nucleotide			
WP	Sequence Split Into 7	Fragments	LOCUS ACF65387 Accession ACF65387
WP	Fragment Name	Begin	End
WP	ACF65387_0	1	110000
WP	ACF65387_1	100001	210000
WP	ACF65387_2	200001	310000
WP	ACF65387_3	300001	410000
WP	ACF65387_4	400001	510000
WP	ACF65387_5	500001	610000
WP	ACF65387_6	600001	696798



## Alignment Scores:

Pred. No.: 0.00125 Length: 110000  
Score: 120.00 Matches: 25  
Percent Similarity: 55.06% Conservative: 24  
Best Local Similarity: 28.09% Mismatches: 40  
Query Match: 23.35% Indels: 0  
DB: 10 Gaps: 0

US-10-089-514-4 (1-103) x ACF65387\_1 (1-110000)

QY 1 MetThrGluGlnAengLueGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20  
DB 13584 ATAAATTAACGGAATATTATACACCTTACGAGAACTTGTGATTATTAATATGACG 13643  
QY 21 LeuLeuAspThrValArgArgArgIleAspLeuGlyValArgIleAlaArgTyrLysSer 40  
DB 13644 ATTTTGGCACTGCTTTCTGAACGATGAAAGTTGTATGAAAACGCTGAGGTAAAGCG 13703  
QY 41 ArgHisGlyValPrometMetGlnProGlyArgValSerLeuValLysAspArgAlaAla 60  
DB 13704 GGGCAGATATTCGATGATGACGACCAACGATTTACCTGTTATTTGAATATGTCGCG 13763  
QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIle 80  
DB 13764 GATTAATCAACCGATTTGGGTTACGGCTGAATATATCTGAATCAATATTTCACTGCTT 13823  
QY 81 IleThrGluMetCysArgValGluAsp 89  
DB 13824 ATTGAAGAGACTTGCCGTCGAGAGAA 13850

## RESULT 10

Continuation (13 of 17) of AAV21209 from base 1200001 (Methanococcus jannaschii circular  
WP Sequence split into 17 fragments LOCUS AAV21209 Accession AAV21209

WP	Fragment Name	Begin	End
WP	AAV21209_00	1	110000
WP	AAV21209_01	100001	210000
WP	AAV21209_02	200001	310000
WP	AAV21209_03	300001	410000
WP	AAV21209_04	400001	510000
WP	AAV21209_05	500001	610000
WP	AAV21209_06	600001	710000
WP	AAV21209_07	700001	810000
WP	AAV21209_08	800001	910000
WP	AAV21209_09	900001	1010000
WP	AAV21209_10	1000001	1110000
WP	AAV21209_11	1100001	1210000
WP	AAV21209_12	1200001	1310000
WP	AAV21209_13	1300001	1410000
WP	AAV21209_14	1400001	1510000
WP	AAV21209_15	1500001	1610000
WP	AAV21209_16	1600001	1664976

## Alignment Scores:

Pred. No.: 0.0158 Length: 110000  
Score: 112.00 Matches: 20  
Percent Similarity: 53.09% Conservative: 23  
Best Local Similarity: 24.69% Mismatches: 38  
Query Match: 21.79% Indels: 0  
DB: 2 Gaps: 0

US-10-089-514-4 (1-103) x AAV21209\_12 (1-110000)

QY 1 MetThrGluGlnAengLueGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20  
DB 38358 GTTACATAGATAGAACTTCTGTAATATGAGAAAGATTGATGATTAATGACATAG 38417  
QY 21 LeuLeuAspThrValArgArgArgIleAspLeuGlyValArgIleAlaArgTyrLysSer 40  
DB 38418 ATATTTAAGCTAATATGCTGAAGAAATAGTTTACTTAAGATGAGCTGAGATTAATAAT 38477  
QY 41 ArgHisGlyValPrometMetGlnProGlyArgValSerLeuValLysAspArgAlaAla 60

DB 38478 CAGCTTGATTCCTATTAAAGACCCAGAAAGAAAATATATATATACATAGATATAG 38537

QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIle 80

DB 38538 AAACCTTTGTAAGAACATACATGTTATGAAAATATTTGCATTAATAATTTCAAAATCTT 38597

QY 81 Ile 81

DB 38598 ATA 38600

## RESULT 11

ADN27081\_16/c  
Continuation (17 of 17) of ADN27081 from base 1600001 (Hyperthermophile Methanopyrus ka

WP	Fragment Name	Begin	End
WP	ADN27081_00	1	110000
WP	ADN27081_01	100001	210000
WP	ADN27081_02	200001	310000
WP	ADN27081_03	300001	410000
WP	ADN27081_04	400001	510000
WP	ADN27081_05	500001	610000
WP	ADN27081_06	600001	710000
WP	ADN27081_07	700001	810000
WP	ADN27081_08	800001	910000
WP	ADN27081_09	900001	1010000
WP	ADN27081_10	1000001	1110000
WP	ADN27081_11	1100001	1210000
WP	ADN27081_12	1200001	1310000
WP	ADN27081_13	1300001	1410000
WP	ADN27081_14	1400001	1510000
WP	ADN27081_15	1500001	1610000
WP	ADN27081_16	1600001	1694968

## Alignment Scores:

Pred. No.: 0.0333 Length: 94968  
Score: 109.00 Matches: 25  
Percent Similarity: 53.01% Conservative: 19  
Best Local Similarity: 30.12% Mismatches: 35  
Query Match: 21.21% Indels: 4  
DB: 1 Gaps: 1

US-10-089-514-4 (1-103) x ADN27081\_16 (1-94968)

QY 5 AsnGluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThr 24  
DB 17909 AACGTGCTCGAGAGCTCAGCGCCGGAATATGACCGCATGATGATTTGCTGAGACGA 17850  
QY 25 ValArgArgArgIleAspLeuGlyValArgIleAlaArgTyrLysSerArgHisGlyVal 44  
DB 17849 GTCATCGAGCGCTCTGAAGGTGCGCCGGAATATGGCGGCTTAAGGCTCAGAGGGGTTA 17790  
QY 45 PrometMetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyrAlaAla 64  
DB 17789 CCACTGACGAGCAGAGGCGGAGAGAGAGCTCCGGAGAGCTGAGAGAGGTTTAA 17730  
QY 65 AspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIleThrGluMet 84  
DB 17729 ACAGAGGGGCTGATCCGGC-----CTCGCGGACATCGTGGCTTCCTATC 17682  
QY 85 CysArgVal 87  
DB 17681 CTAAAGGTT 17673  
RESULT 12  
ADP00853  
ID ADP00853 standard; DNA; 1125 BP.  
AC ADP00853;  
XX  
XX 12-FEB-2004 (first entry)  
XX  
XX Bacterial polynucleotide #1138.  
XX



KW Proteus mirabilis infection; bacterial infection; antibacterial;  
KW immunostimulant; gene; ds.  
XX  
OS Proteus mirabilis.  
XX  
PN US6605709-B1.  
XX  
PD 12-AUG-2003.  
XX  
PF 05-APR-2000; 2000US-00543681.  
XX  
PR 09-APR-1999; 99US-0128706P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Breton GL;  
XX  
DR WPI; 2003-895291/82.  
DR P-PSDB; ADF05025.  
XX  
XX New Proteus mirabilis polypeptides and polynucleotides, useful as  
PT reagents for diagnosis of bacterial disease, as components of  
PT antibacterial vaccines, as targets for antibacterial drugs, or as  
PT biocontrol agents for plants.  
XX  
PS Disclosure; SEQ ID NO 1138; 870pp; English.  
XX  
XX The invention relates to new Proteus mirabilis polypeptides and  
CC polynucleotides. The invention also relates to antibodies against the  
CC polypeptides, methods for producing the polypeptides, a method of  
CC generating vaccines for immunising an individual against P. mirabilis, a  
CC method for evaluating a compound for the ability to bind a P. mirabilis  
CC polypeptide and a method for screening test compounds for anti-bacterial  
CC activity. The polypeptides and polynucleotides are useful as molecular  
CC targets for diagnosing, preventing and treating pathological conditions  
CC resulting from bacterial infection, as reagents for diagnosis of  
CC bacterial diseases, as components of antibacterial vaccines, as targets  
CC for antibacterial drugs or as bio-control agents for plants. This  
CC sequence represents a Proteus mirabilis polynucleotide of the invention.  
XX  
SQ Sequence 1125 BP; 352 A; 180 C; 250 G; 343 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 0.000197 Length: 1125  
Score: 106.00 Matches: 27  
Percent Similarity: 46.94% Conservative: 19  
Best Local Similarity: 27.55% Mismatches: 36  
Query Match: 20.62% Indels: 16  
DB: 10 Gaps: 2  
  
US-10-089-514-4 (1-103) x ADF00853 (1-1125)  
  
QY 6 GlnLeuGlnArgLeuArgAlaGlnLeuAspAlaLeuAspGlyThrLeuLeuAspThrVal 25  
DB 10 GAATTTCACAAATTTGGAGAGCAAAATTGATCAGTGATTAACATTTCTTTTCACTC 69  
QY 26 ArgArgArgIleAspLeuGlyValArgIleAlaArgTyrIleSerArgHisGlyValPro 45  
DB 70 GCCAAAGAAATGCAATTTGGTAGCAAGATGGGAAAGTTAAATAATGCGCGGGTTACCT 129  
QY 46 MetMetGlnProGlyArgValSerLeuValIleAspArgAlaAlaArgTyrAlaAlaAsp 65  
DB 130 ATGTATGACCAACATATAGAGAAGCGCAATCTTCTTCGAGCTATAGACGAAAGAAA 189  
QY 66 HisGlyLeuAspGlnSerPheLeuValAsnLeuTyrAspValIleIleThrGlnMetCys 85  
DB 190 ATGGGATTTCTCTGCTGATTTA-----MetSerArgGlnSerLeuThrAlaGlnAsp 210  
QY 86 ArgValGlnAspLeuVal-----MetSerArgGlnSerLeuThrAlaGlnAsp 101  
DB 211 ---ATTGAAGATGTTTAACTGTCGAGTAAATGAGAGTCAATATAGCAAAAGAAAT 261  
RESULT 13

ACH98708  
ID ACH98708 standard; DNA; 1305 BP.  
XX  
XX ACH98708;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Klebsiella pneumoniae polynucleotide seqid 4503.  
XX  
XX  
XX Recombinant expression vector; transcription regulatory element;  
KW Klebsiella pneumoniae protein; antibacterial; vaccine; gene; ds.  
XX  
OS Klebsiella pneumoniae.  
XX  
PN US6610836-B1.  
XX  
PD 26-AUG-2003.  
XX  
PF 27-JAN-2000; 2000US-00489039.  
XX  
PR 29-JAN-1999; 99US-0117747P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Breton GL, Osborne M;  
XX  
DR WPI; 2003-895346/82.  
DR P-PSDB; ABO65157.  
XX  
XX New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for  
PT preparing a vaccine composition against Klebsiella pneumoniae.  
XX  
PS Disclosure; SEQ ID NO 4503; 932pp; English.  
XX  
XX The invention describes a new isolated nucleic acid encoding a Klebsiella  
CC pneumoniae polypeptide. Also described are: a recombinant expression  
CC vector comprising the nucleic acid, operably linked to a transcription  
CC regulatory element; and a cell comprising the recombinant expression  
CC vector. The nucleic acid is useful for preparing a vaccine composition  
CC against Klebsiella pneumoniae. This sequence encodes a Klebsiella  
CC pneumoniae polypeptide of the invention  
XX  
SQ Sequence 1305 BP; 290 A; 388 C; 360 G; 267 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 0.000332 Length: 1305  
Score: 105.00 Matches: 29  
Percent Similarity: 48.45% Conservative: 18  
Best Local Similarity: 29.90% Mismatches: 44  
Query Match: 20.43% Indels: 6  
DB: 11 Gaps: 1  
  
US-10-089-514-4 (1-103) x ACH98708 (1-1305)  
  
QY 1 MetThrGlnGlnAsnGlnLeuGlnArgLeuArgAlaGlnLeuAspAlaLeuAspGlyThr 20  
DB 145 ATGACCGAGAGAAACCATTAATCTGCGCGCTGCCGATTAATAGAGCCCTGAGCAGGAAG 204  
QY 21 LeuLeuAspThrValArgArgArgIleAspLeuGlyValArgIleAlaArgTyrIleSer 40  
DB 205 CTGCTTGCTCTCTCGCGAGCGCGCGGATGCTGTTGAGGTTGGAAGAACCAAACTG 264  
QY 41 ArgHisGlyValProMetMetGlnProGlyArgValSerLeuValIleAspArgAlaAla 60  
DB 265 GCCTCCCATGTCGGTGGTGATATTGACCGCGAAGCGGATCTGCTGAGGCGCTGATG 324  
QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGlnSerPheLeuValAsnLeuTyrAspValIle 80  
DB 325 ACTATCGGACGAGCGCATATATCTGATGCGCACTATATATACCGGCTGTCCAGCTGATT 384  
QY 81 IleThrGlnMetCysArgValGlnAspLeuValMetSerArgGlnSerLeu 97  
DB 385 ATC-----GAGACTCTGTTCTTACCCAGCAAAACCTG 417











GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 10:07:08 : Search time 56.3672 Seconds  
(without alignments)  
2989.973 Million cell updates/sec

Title: US-10-089-514-4

Perfect score: 514  
Sequence: 1 MTEQNELQLRAELDALDGT.....MCRVEDLVMSRESLTAEDDR 103

## Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000Post-Processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Command line parameters:

-MODE=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/uspto.spool\_h/US10089514/runat\_04102005\_105746\_8071/app\_query.fasta\_1.1621  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10089514\_@CGN\_1\_116\_@runat\_04102005\_105746\_8071 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

Issued Patents NA.\*  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/prodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/prodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	256	49.8	387	3	US-08-765-907A-4
2	256	49.8	387	3	US-09-987-614A-4
3	256	49.8	2888	3	US-08-765-907A-1
4	256	49.8	2888	4	US-09-987-614A-1
5	112	21.8	1664976	4	US-08-916-421B-1
6	112	21.8	1664976	4	US-09-692-570-1
7	106	20.6	1135	4	US-09-543-681A-1138
8	105	20.4	1305	4	US-09-489-039A-4503
9	104.5	20.3	15639	4	US-09-902-540-2740
10	104.5	20.3	15639	4	US-09-902-540-1129
11	98	19.1	850	4	US-09-221-017B-361
12	93	18.1	1131	4	US-09-489-039A-4572

13	90	17.5	1038	4	US-09-902-540-5637	Sequence 5637, Ap
14	90	17.5	49225	4	US-09-902-540-1269	Sequence 1269, Ap
15	88.5	17.2	804	4	US-09-252-991A-4814	Sequence 4814, Ap
16	88.5	17.2	960	4	US-09-252-991A-4705	Sequence 4705, Ap
17	88.5	17.2	1653	4	US-09-252-991A-4724	Sequence 4724, Ap
18	88.5	17.2	1791	4	US-09-252-991A-4684	Sequence 4684, Ap
19	88.5	17.2	1830121	4	US-09-557-884-1	Sequence 1, Appli
20	88.5	17.2	1830121	4	US-09-643-990A-1	Sequence 1, Appli
21	82	16.0	1167	4	US-09-543-681A-1120	Sequence 1320, Ap
22	76	14.8	267	4	US-09-583-110-1836	Sequence 1836, Ap
23	75	14.6	4258	3	US-07-765-830A-5	Sequence 5, Appli
24	75	14.6	4825	4	US-09-949-016-15912	Sequence 15912, A
25	74	14.4	888	4	US-09-252-991A-5830	Sequence 5830, Ap
26	74	14.4	1041	4	US-09-252-991A-5815	Sequence 5815, Ap
27	74	14.4	1066	4	US-09-252-991A-10259	Sequence 10259, A
28	74	14.4	1101	4	US-09-252-991A-10597	Sequence 10597, A
29	74	14.4	1581	4	US-09-252-991A-10473	Sequence 10473, A
30	74	14.4	1611	4	US-09-252-991A-5778	Sequence 5778, Ap
31	74	14.4	9839	4	US-09-902-540-996	Sequence 996, App
32	73	14.2	921	4	US-09-902-540-7847	Sequence 7847, Ap
33	73	14.2	5125	4	US-09-902-540-785	Sequence 785, App
34	72	14.0	447	4	US-09-902-540-9336	Sequence 9326, Ap
35	71.5	13.9	1230	4	US-10-105-319-1	Sequence 1, Appli
36	71.5	13.9	1281	3	US-09-105-537-19	Sequence 19, Appli
37	71.5	13.9	13613	3	US-09-105-537-3	Sequence 3, Appli
38	71.5	13.9	18431	3	US-09-221-017B-1090	Sequence 1090, Ap
39	71.5	13.9	38506	3	US-09-320-878-19	Sequence 19, Appli
40	71.5	13.9	38506	4	US-09-141-908-1	Sequence 1, Appli
41	71.5	13.9	38506	4	US-09-657-440-19	Sequence 19, Appli
42	71	13.8	1915	4	US-09-902-540-7256	Sequence 7256, Ap
43	71	13.8	6744	4	US-09-902-540-682	Sequence 682, App
44	69.5	13.5	1524	4	US-09-540-236-1471	Sequence 1471, Ap
45	69	13.4	1125	4	US-09-902-540-8770	Sequence 8770, Ap

## ALIGNMENTS

RESULT 1  
US-08-765-907A-4  
Sequence 4, Application US/08765907A  
Patent No. 6352839  
GENERAL INFORMATION:  
APPLICANT: BLANC, Veronique  
APPLICANT: THIBAUT, Denis  
APPLICANT: BMMAS-JACQUES, Nathalie  
APPLICANT: BLANCHE, Francis  
APPLICANT: COUZET, Joel  
APPLICANT: BARRIERE, Jean-Claude  
APPLICANT: DEBUSSCHE, Laurent  
APPLICANT: PAMECHON, Alain  
APPLICANT: PARIS, Jean-Marc  
APPLICANT: DUTROC-ROSSET, Gilles  
TITLE OF INVENTION: Streptogramins And Method For Preparing Same By  
FILE REFERENCE: Streptogramin genes  
CURRENT APPLICATION NUMBER: US/08/765,907A  
CURRENT FILING DATE: 1997-03-20  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 387  
TYPE: DNA  
ORGANISM: Streptomyces pristinaespiralis  
US-08-765-907A-4

## Alignment Scores:

Pred. No.: 4,96e-28  
Score: 256.00  
Percent Similarity: 71.08%  
Best Local Similarity: 61.45%  
Query Match: 49.81%  
DB: 3  
Length: 387  
Matches: 51  
Conservative: 8  
Mismatch: 24  
Indels: 0  
Gaps: 0







```

; APPLICANT: THIBAUT, Denis
; APPLICANT: BAWAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSCHÉ, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSER, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US/09/987, 614A
; PRIOR FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2888
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-1

Alignment Scores:
Pred. No.:      8,62e-27      Length:      2888
Score:          256.00        Matches:      51
Percent Similarity: 71.08%    Conservative: 8
Best Local Similarity: 61.45%  Mismatches:  24
Query Match:     49.81%      Indels:       0
DB:              4           Gaps:          0

US-10-089-514-4 (1-103) x US-09-987-614A-1 (1-2888)

QY      7  LeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrValArg 26
Db      1936  CTCGACCGCTGCGCGCGCGCGCTCGACCGCGCGCGCTGCGACCGCGCTGCGACCGCGCTCGC 1995

QY      27  ArgArgGLeuAspLeuGlyValArgGLeuAlaArgGlyTyrSerArgGHisGlyValPromet 46
Db      1996  ACACGCGCTGACATCTCGCTCGGATCGGCGAGTACAAAGCCCTCCACCAAGTCCCATG 2055

QY      47  MetGlnProGlyArgValSerLeuValAsnLeuAspArgAlaAlaArgTyrAlaAlaAspHis 66
Db      2056  ATGACGCGCCACCGCGGATCGCCAGGTCCACCGCCCAAGCCCGCCCTACCGCCGACAC 2115

QY      67  GlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIleThrGluMetCysArg 86
Db      2116  GGCATCGACCGCGCGCTTCTGCGCAGCCCTGTACACAGATCATCACCGAGACTGCGCGC 2175

QY      87  ValGluAsp 89
Db      2176  CTCGAGGAC 2184

RESULT 5
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Built et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6503729
; TITLE OF INVENTION: Jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916, 421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024, 428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
```

```

ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (10398)..(10398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231580)..(231580)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226)..(319226)
```



```
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
```

```
NAME/KEY: misc_feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1664854)..(1664854)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Alignment Scores:
Pred. No.: 0.0612 Length: 1664976
Score: 112.00 Matches: 20
Percent Similarity: 53.09% Conservative: 23
Best Local Similarity: 24.69% Mismatches: 38
Query Match: 21.79% Indels: 0
DB: 4 Gaps: 0

US-10-089-514-4 (1-103) x US-08-916-421B-1 (1-1664976)

QY 1 MetThrGlulnAsnGlulnArgLeuArgAlaGlulnAspAlaLeuAspGlyThr 20
Db 1238358 GTTACATGATAGAGAACTTCCTGAAATTAGGAAAGATGATGACATTAAG 1238417
QY 21 LeuLeuAspThrValArgArgArgIleAspLeuGlyValArgIleAlaArgTyrIleSer 40
Db 1238418 ATATTAAAGCTTAATGCTGAAAGAAATTAAGTCTTAAGCATGATGCTAGATTAATAAT 1238477
QY 41 ArgHisGlyValProMetMetGlnProGlyArgValSerLeuValIleAspArgAlaIle 60
Db 1238478 CAGCTTGATTCCTATTTAACGACCCAGAAAGAAAAATATATATACATGATTAAGA 1238537
QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGlySerPheLeuValAsnLeuTyrIle 80
Db 1238538 AAACCTTGTAAGAAACATTAACGTTGTAATAATTTGCATTAATAATTCATAATACCTT 1238597
QY 81 Ile 81
Db 1238598 ATA 1238600

RESULT 6
US-09-692-570-1
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
; APPLICANT: Buit et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6797466
; TITLE OF INVENTION: jannaschi
; FILE REFERENCE: PB275CI
; CURRENT APPLICATION NUMBER: US/09/692,570
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; PRIOR APPLICATION NUMBER: US 08/916,421
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschi
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28232)..(28232)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
```



OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (84773)..(84773)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (84808)..(84808)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (84812)..(84812)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (98120)..(98120)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (98159)..(98159)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (98239)..(98239)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (98266)..(98266)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (98343)..(98343)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (103998)..(103998)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (148948)..(148948)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (163385)..(163385)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (191989)..(191989)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (191995)..(191995)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (231980)..(231980)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (234187)..(234187)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (234220)..(234220)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (234814)..(234814)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (309398)..(309398)  
OTHER INFORMATION: n equals a, t, c, or g

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (309418)..(309418)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (312837)..(312837)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (312993)..(312993)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (319226)..(319226)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (559167)..(559167)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (559241)..(559241)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (600992)..(600992)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (622708)..(622708)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (657081)..(657081)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (657203)..(657203)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (674435)..(674435)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (682442)..(682442)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (713652)..(713652)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (741684)..(741684)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (779455)..(779455)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (779676)..(779676)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (855539)..(855539)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (871619)..(871619)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:







```

Oy      41  ATGHSGLGValPProMetGlnPProGlyArgValSerLeuVallyAspArgAla 60
Db      265  GGCTCCCACTGTCGGTGGCGTATATTAGACCGGAACGGCATCTGCTGGAGCCCTGATG 324
Oy      61  ATGTYrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTYrAspValIle 80
Db      325  ACTATGCGCAACGGCATATATCTGATGCGACATATATACCCGCGCTGTTCCAGCTGATT 384
Oy      81  ILeThrGluMetCysArgValGluAspLeuValMetSerArgGluSerLeu 97
Db      385  ATC-----GAAAGACTGTCGTCTTACCAACGAAACCCCTG 417

RESULT 9
US-09-902-540-2740
; Sequence 2740, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Steven C.
; APPLICANT: Slater, Gregory J.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 2740
; LENGTH: 1963
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-2740

Alignment Scores:
Pred. No.:      5, 2e-05      Length:      1963
Score:          104.50      Matches:      28
Percent Similarity: 48.91%      Conservative: 17
Best Local Similarity: 30.43%      Mismatches: 46
Query Match:      20.33%      Indels:      1
DB:              4          Gaps:      1

US-10-089-514-4 (1-103) x US-09-902-540-2740 (1-1963)
Oy      1  MetThrGluGlnAsnGluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20
Db      1  ATGACCAACCCTCCCGGACCTCGACCTGCATCCGACCTCCATCGAGGCGCATCGACGAGAG 60
Oy      21  LeuLeuAspThrValAlaArgArgArgIleAspLeuGlyValAlaArgIleAlaArgTYrLysSer 40
Db      61  ATCTTCGARGCCCTCCGCGCGGAGATGGCGCTGGCGGACGACGACCTGCGCCGCGGAGACTG 120
Oy      41  ARGHSGLGValPProMetMetGlnPProGlyArgValSerLeuVallyAspArgAlaAla 60
Db      121  GCCGCGCGCGCGCCCTTCGCGGAGCACCGCGCGGAGAGACTGCTCTCGCGCGGATCCGC 180
Oy      61  ATGTYrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTYrAspValIle 80
Db      181  ACCCGCGCGCGCGGAGACGCGCTCGAGACCCGCAAGAGTGAAGCGCATCTGGCGGCTTTC 240
Oy      81  ILe---ThrGluMetCysArgValGluAspLeuVal 91
Db      241  ATGACATGTCTCGTGGCTCGGCAACACGAGCTGCT 276

RESULT 10
US-09-902-540-1129/c
; Sequence 1129, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.

```

```

; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(1)58491B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1129
; LENGTH: 15689
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1129

Alignment Scores:
Pred. No.:      0.000997      Length:      15689
Score:          104.50        Matches:     28
Percent Similarity: 48.91%    Conservative: 17
Best Local Similarity: 30.43% Mismatches:   46
Query Match:    20.33%       Indels:      1
DB:              4           Gaps:         1

US-10-089-514-4 (1-103) x US-09-902-540-1129 (1-15689)

QY      1 MetThrGluGlnAengIuleGlnArgLeuAlaGluLeuAspAlaLeuAspGlyThr 20
Dd      8044 ATGCACCACTCCGCCCGAGCTTGACTTATCCGGACTTCATTGAGCCCATGACAGAG 7985
QY      21 LeuLeuAspThrValAlaArgArgArgIleAspLeuGlyValArgIleAlaArgTYrLysSer 40
Dd      7984 ATCTTCGATGCGCTCCGCCCGCGGATGCGGTGGCGGACAGACGTCGCCGCCGCGGACGCTG 7925
QY      41 ArgHisGlyValProMetMetGlnProGlyArgValSerLeuValLysAspArgAlaAla 60
Dd      7924 GCCCGCCCCCGCGCCTTCCGGGACCAGCGCGCGGAGGACCTGCTCCGCCGCCCATCCGC 7865
QY      61 ArgTYrValAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTYrAspValIle 80
Dd      7864 ACCCGCGCGCGGAGCACGCGCTGAGCCCGACGAGGTGAACCGCATCTGAGGCTCTTC 7805
QY      81 Ile---ThrglMetCysArgValGluAspLeuVal 91
Dd      7804 ATCGACATGTCCTCGGCTCGGCAACAGACTCTCTC 7769

RESULT 11
US-09-221-017B-361/C
; Sequence 361, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA: PP1546
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998

```



```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 361:
SEQUENCE CHARACTERISTICS:
LENGTH: 850 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORYPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...850
US-09-221-017B-361

Alignment Scores:
Pred. No.: 0.00014 Length: 850
Score: 98.00 Matches: 21
Percent Similarity: 56.34% Conservative: 19
Best Local Similarity: 29.58% Mismatches: 31
Query Match: 19.07% Indels: 0
DB: 3 Gaps: 0

US-10-089-514-4 (1-103) x US-09-221-017B-361 (1-850)
QY 3 GluGlnaAngLueGluGlnArgLeuArgLagLueAspAlaLeuaspLcylThrLeuLeu 22
   :: ||||| ||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
Db 214 CAAAGCGAAGACTATCTCTGGCGCATGCAGATTGATCAATGATGAGGTATAGTG 155
QY 23 AspThrValaArgArgArgIleaspLeuGlyValaArgIlealArgTyrIysSerArgHis 42
   :: ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db 154 GAATTGCTAGCTCGCGCGATGCGAATGGCATGCGAATGCGTTGTTCAAAAAAGAGCAC 95
QY 43 GlyValPProMetMetGlnProGlyArgValSerLeuValLysaspArgAlaIalArgTyr 62
   :: ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db 94 AATCGGCTGTGGTTCAGAAATCTCGCTACGAAACAACACGCAACCGTCCCGTACT 35
QY 63 AlaAlaaspHisGlyLeuaspGluSerPheLeu 73
   ||||| ||||| ||||| ||||| ::||| ::||| ::||| ::|||
Db 34 GAGGCGCTTAGGTGGAGCAAAACATTATATA 2

RESULT 12
US-09-489-039A-4572
Sequence 4572, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709, 2004001
CURRENT APPLICATION NUMBER: US/09/489, 039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 4572
LENGTH: 1131
TYPE: DNA

```

```

; ORGANISM Klebsiella pneumoniae
US-09-489-039A-4572

Alignment Scores:
Pred. No.: 0.00112 Length: 1131
Score: 93.00 Matches: 26
Percent Similarity: 47.96% Conservative: 21
Best Local Similarity: 26.53% Mismatches: 35
Query Match: 18.09% Indels: 16
DB: 4 Gaps: 2

US-10-089-514-4 (1-103) x US-09-489-039A-4572 (1-1131)

QY      6  G|LueGlnArgLeuValArgAlaIleuLeuAspAlaLeuAspGlyThrLeuLeuAspThrVal 25
        |||||
Db       19  GAACGTAGCCGGCTTACGCATCAAAATTGATGAAGTGATAAAGCGTTGCTTAGGCCCTG 78
              :::::

QY     26  ArgArgGluLeuAspLeuGlyValArgIleAlaArgTrlyrSerArgHISglValPro 45
        :::::
Db      79  GCTTAAGGTTGAGTCGTGTGGCCCGAGAAGTCGGCAGGTGAAGAGCCAGTATGGCTGCG 138
              :::::

QY     46  MetMetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTYRAlaAlaAsp 65
        :|::|
Db     139  ATTTCAGTCCCGGAGCGCAGTCGCGCATGCTCGCGCTCCGACGTGAAGAGCGCGCTCG 198
              :::::

QY     66  HISgLyLeuAspGluSerPheLeuValAsnLeuTYRAspValILElleThrglumeCYs 85
        :::::
Db     199  CTCGGCGTACCACCGCATCTG-----MetSerArgGluSerLeuThrAlaGluasp 219
              :::::

QY     86  ArgValGluAspLeuVal-----MetSerArgGluSerLeuThrAlaGluasp 101
        :|::|
Db     220  --ATCGAAGACGTTCTGTGCGTCGCGTGAATGCGGGAATCTATTCACGAAAAAC 270
              :::::

RESULT 13
US-09-902-540-5637
; Sequence 5637, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 5637
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-5637

Alignment Scores:
Pred. No.: 0.0027 Length: 1038
Score: 90.00 Matches: 20
Percent Similarity: 53.33% Conservative: 20
Best Local Similarity: 26.67% Mismatches: 35
Query Match: 17.51% Indels: 0
DB: 4 Gaps: 0

US-10-089-514-4 (1-103) x US-09-902-540-5637 (1-1038)

QY      7  LeuGlnArgLeuValArgAlaIleuLeuAspAlaLeuAspGlyThrLeuLeuAspThrValArg 26
        ::|||
Db      775  CTGACCGAAGTTGCGGAGTGCGTGATCGTTGACCGTCGATGCTTCAACTCCGTAAT 834
              :::::

QY     27  ArgArgGluLeuGlyValArgIleAlaArgTrlyrSerArgHISglValPromet 46
        ::|||
Db     835  CGTGGCGCCAGCTATTCACGAGGCCACACCTGAAGGCCGAGCAGGACTTCGCGCTG 894
              :::::

```



```
QY 47 MetGlnProGlyArgValSerLeuValAspArgAlaAlaArgTyrAlaAlaAspHis 66
DB 895 CCAGACCGCGGAGCGCGAGCGAGCTCTGTGAGACCCGGCGGTGCGCCGCGAGCAG 954
QY 67 GlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIle 81
DB 955 GGCATGACCGCGAGCAGACGAAAGAGTGTTCGCCGCGCTCTG 999

RESULT 14
US-09-902-540-1269/C
; Sequence 1269, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1269
; LENGTH: 49225
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1269

Alignment Scores:
Pred. No.: 0 648 Length: 49225
Score: 90.00 Matches: 20
Percent Similarity: 53.33% Conservative: 20
Best Local Similarity: 26.67% Mismatches: 35
Query Match: 17.51% Indels: 0
DB: 4 Gaps: 0

US-10-089-514-4 (1-103) x US-09-902-540-1269 (1-49225)
QY 7 LeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrValArg 26
DB 40904 CTGACCGAGAGTTCGCGAGCGAGCTGTGAGACCGTGTGACCTGTCAACTCTGAAT 40845
QY 27 ArgArgIleAspLeuGlyValArgIleAlaArgTyrLysSerArgHisGlyValProMet 46
DB 40844 CGTCGCCGCCAGCTCATCCAGCAGCCGACACCTGAAGGCCGAGCAGGACTTCCGCTG 40785
QY 47 MetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyrAlaAlaAspHis 66
DB 40784 CCAGACCGCGGAGCGCGAGCGAGCTCTGTGAGACCCGGCGGTGAGCCGCGAGCAG 40725
QY 67 GlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIle 81
DB 40724 GGCATGACCGCGAGCAGACGAAAGAGTGTTCGCCGCGCTCTG 40680

RESULT 15
US-09-252-991A-4814/C
; Sequence 4814, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 31142
; SEQ ID NO 4814
```

```
; LENGTH: 804
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4814

Alignment Scores:
Pred. No.: 0 0031 Length: 804
Score: 88.50 Matches: 24
Percent Similarity: 53.95% Conservative: 17
Best Local Similarity: 31.58% Mismatches: 32
Query Match: 17.22% Indels: 3
DB: 4 Gaps: 2

US-10-089-514-4 (1-103) x US-09-252-991A-4814 (1-804)
QY 7 LeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrValArg 26
DB 690 CTGCGGAGACTTCGCGAGCGAGCATCGACCGATGACCTGGATTCGTCAGGCCCTCGGC 631
QY 27 ArgArgIleAspLeuGlyValArgIleAlaArgTyrLysSerArgHisGlyValProMet 45
DB 630 CGCCGATGAGACTCACTCAAGCGCGCGTCCGCTTCAAGCCAGCAGCGCGGATTCG 571
QY 46 MetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyrAlaAlaAsp 65
DB 570 -----GCACCGAGCGGGTCCGCGGATGCTCCCGAGCGCGCCGCTGCGCGAGAA 517
QY 66 HisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIle 81
DB 516 AACGAGTGCAGCGCGCTTGTGAGGAGACTGTTCGCGGATCATC 469

Search completed: October 5, 2005, 21:11:55
Job time : 605.367 secs
```



This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 09:57:48 ; Search time 226.674 Seconds

(without alignments)  
3137.453 Million cell updates/sec

Title: US-10-089-514-4

Perfect score: 514

Sequence: 1 MTEQNELQRRARLALDGT.....MCRVEDLVMSRESLTAEDDR 103

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 7442561 seqs, 3452328358 residues

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing:

Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cg2\_1/USPTO.spool.h/US10089514/runat\_04102005\_105745\_8055/app\_query.fasta\_1.1621  
-DB=published Applications NA -QFMT=fastab -SUFFIX=tmp -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blonum62  
-TRANS=human40.cdt -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptco -NORM=ext -HAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10089514@cgn\_1\_1\_920@runat\_04102005\_105745\_8055  
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published\_Applications\_NA:\*

1: /cg2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cg2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cg2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cg2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cg2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cg2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cg2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cg2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cg2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cg2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cg2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cg2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cg2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
14: /cg2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
15: /cg2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*  
16: /cg2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*  
17: /cg2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*  
18: /cg2\_6/ptodata/1/pubpna/US10G\_PUBCOMB.seq:\*  
19: /cg2\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq:\*  
20: /cg2\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq:\*  
21: /cg2\_6/ptodata/1/pubpna/US10J\_PUBCOMB.seq:\*  
22: /cg2\_6/ptodata/1/pubpna/US10K\_PUBCOMB.seq:\*  
23: /cg2\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq:\*  
24: /cg2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*  
25: /cg2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
26: /cg2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB ID	Description
1	514	100.0	312	20	US-10-472-587-3
2	486.5	94.6	3305	14	US-10-126-927-68
3	439	85.4	12391	14	US-10-126-927-67
4	102	19.8	1122	16	US-10-137-310-1
5	102	19.8	1122	18	US-10-634-548-16
6	102	19.8	1152	19	US-10-433-556-20
7	98	19.1	850	13	US-10-194-163-361
8	96.5	18.8	1833	17	US-10-381-779-40
9	91	17.7	387	15	US-10-156-761-6187
10	91	17.7	9025608	15	US-10-156-761-1
11	90	17.5	1132	16	US-10-137-310-3
12	90	17.5	1132	19	US-10-433-556-18
13	90	17.5	1238	16	US-10-380-132-7
14	89.5	17.4	1104	17	US-10-282-122A-15346
15	88.5	17.2	1830121	17	US-10-329-670-1
16	88.5	17.2	1830121	20	US-10-158-865-1
17	88.5	17.2	1830121	22	US-10-981-687-1
18	86	16.7	951	15	US-10-156-761-2533
19	82.5	16.1	300	18	US-10-335-977-4562
20	82.5	16.1	303	18	US-10-335-977-4561
21	78.5	15.3	1868	13	US-10-027-632-99311
22	78.5	15.3	1868	13	US-10-027-632-99312
23	78.5	15.3	1868	17	US-10-027-632-99311
24	78.5	15.3	1868	17	US-10-027-632-99312
25	78.5	15.3	1869	13	US-10-027-632-98019
26	78.5	15.3	1869	17	US-10-027-632-98019
27	77.5	15.0	1920	22	US-10-450-763-30061
28	77	15.0	1920	22	US-10-472-928-2611
29	75	14.6	224	21	US-10-027-632-166491
30	75	14.6	821	13	US-10-027-632-166491
31	75	14.6	1121	17	US-10-369-493-24241
32	75	14.6	1121	17	US-10-369-493-24241
33	75	14.6	16299	15	US-10-102-524-127
34	75	14.6	216298	21	US-10-472-928-4979
35	74	14.4	2282	17	US-10-369-493-36322
36	73	14.2	729	15	US-10-156-761-3396
37	73	14.2	1017	15	US-10-146-772-249
38	73	14.2	1017	17	US-10-241-742-249
39	73	14.2	1017	17	US-10-440-523-249
40	73	14.2	1017	18	US-10-440-503-249
41	73	14.2	1017	18	US-10-461-925-249
42	73	14.2	1882	17	US-10-320-797-2260
43	73	14.2	2089	17	US-10-320-797-1260
44	73	14.2	4089	17	US-10-320-797-260
45	73	14.2	9025608	15	US-10-156-761-1

RESULT 1  
US-10-472-587-3  
Sequence 3, Application US/10472587  
Publication No. US20040214274A1  
GENERAL INFORMATION:  
APPLICANT: YANAI, Naomi  
APPLICANT: SUMIDA, Naomi  
APPLICANT: WATANABE, Manabu  
APPLICANT: MORIYA, Tatsuki  
APPLICANT: MURAKAMI, Takeshi  
TITLE OF INVENTION: Transformants Producing Substance P1022 Derivatives, Methods for  
FILE REFERENCE: 2003-1302A/MMC/00144  
CURRENT APPLICATION NUMBER: US/1010472,587  
CURRENT FILING DATE: 2003-09-22  
PRIOR APPLICATION NUMBER: 82227/2001



```
/ PRIOR FILING DATE: 2001-03-22
/ NUMBER OF SEQ ID NOS: 40
/ SOFTWARE: Patentin Ver. 2.1
/ SEQ ID NO 3
/ LENGTH: 312
/ TYPE: DNA
/ ORGANISM: Streptomyces venezuelae
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(309)
US-10-472-587-3

Alignment Scores:
Pred. No.: 1,26e-68 Length: 312
Score: 514.00 Matches: 103
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 20 Gaps: 0

US-10-089-514-4 (1-103) x US-10-472-587-3 (1-312)

QY 1 MetThGluGlnAsnGluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20
Db 1 ATGACCGAGCAGACGAGCTGACGCGCTGCGCGGAGCTCGACGCGCTCGACGCGGACG 60
QY 21 LeuLeuAspThrValArgArgGlyLeuAspGlyValArgIleAlaArgTyrIlySer 40
Db 61 CTCTGGACACGAGTGGCGCGCGCGCTGACCTCGGTGTCGCGATCGCGCGTACAAATCC 120
QY 41 ArgHisGlyValProMetMetGlnProGlyArgValSerLeuValIysAspArgAla 60
Db 121 CGGACGCGCGCTCCGATGATGACGCGCGCGGCTGAGCTGGTCAAGACAGGCGCGCC 180
QY 61 ArgTyrAlaIleAspHisGlyLeuAspGlySerPheLeuValAsnLeuTyrAspVal 80
Db 181 CGCTACGCGCGCGACCAAGCGCTCGACGAATCGTTCCTGTGAACCTCTACGACGTCATC 240
QY 81 IleThrGluMetCysArgValGluAspLeuValMetSerArgGlySerLeuThrAla 100
Db 241 ATCAACGAGATGTGCGCGCTCGAGACCTGTGATGACCGGAGAGACCTGACGCGCGAG 300
QY 101 AspArgArg 103
Db 301 GACCGGCGG 309

RESULT 2
US-10-126-927-68
/ Sequence 68, Application US/10126927
/ Publication No. US20030082575A1
/ GENERAL INFORMATION:
/ APPLICANT: The Scripps Research Institute
/ APPLICANT: Schultz, Peter G
/ APPLICANT: Wang, Lei
/ APPLICANT: Anderson, John C
/ APPLICANT: Chin, Jason
/ APPLICANT: Liu, David R
/ APPLICANT: Magliery, Thomas
/ APPLICANT: Meggers, Eric L
/ APPLICANT: Mehl, Ryan A
/ APPLICANT: Pasternak, Miro
/ APPLICANT: Santoro, Stephen W
/ APPLICANT: Zhang, Zhiwen
/ TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
/ FILE REFERENCE: 54-000120US
/ CURRENT APPLICATION NUMBER: US/10/126,927
/ CURRENT FILING DATE: 2002-04-19
/ PRIOR APPLICATION NUMBER: US 60/285,030
/ PRIOR FILING DATE: 2001-04-19
/ PRIOR APPLICATION NUMBER: US 60/355,514
/ PRIOR FILING DATE: 2002-02-06
/ NUMBER OF SEQ ID NOS: 79
/ SOFTWARE: Patentin version 3.1
```

```
/ SEQ ID NO 68
/ LENGTH: 3305
/ TYPE: DNA
/ ORGANISM: Streptomyces venezuelae
US-10-126-927-68

Alignment Scores:
Pred. No.: 5,45e-63 Length: 3305
Score: 486.50 Matches: 103
Percent Similarity: 98.10% Conservative: 0
Best Local Similarity: 98.10% Mismatches: 0
Query Match: 94.65% Indels: 2
DB: 14 Gaps: 1

US-10-089-514-4 (1-103) x US-10-126-927-68 (1-3305)

QY 1 MetThGluGlnAsnGluLeuGln---ArgLeuArgAlaGluLeuAspAlaLeuAspG 19
Db 2990 ATGACCGAGCAGAACGAGCTGACGAGTTGGCGCTGCGCGGAGCTCGACGCGTCA 3049
QY 19 YhrLeuLeuAspThrValArgArgGlyLeuAspGlyValArgIleAlaArgTyrIly 39
Db 3050 GACGCTTCTGACACGAGTGGCGCGCGCATGACCTCGGTGTCGCGATGCGCGTACA 3109
QY 39 sSerArgHisGlyValProMetMetGlnProGlyArgValSerLeuValIysAspArgAl 59
Db 3110 GTCCCGGACAGCGCGTCCGATGATGACCGCGCGGCTCAGCTGCTCAAGACAGGCG 3169
QY 59 aAlaArgTyrAlaIleAspHisGlyLeuAspGlySerPheLeuValAsnLeuTyrAspVa 79
Db 3170 CGCCCGCTACGCGCGCGGACACGCGCTCGACGAATCGTTCCTGTGTAACCTCTAC 3229
QY 79 IleIleThrGluMetCysArgValGluAspLeuValMetSerArgGlySerLeuThrAl 99
Db 3230 GATCATCACGAGATGTGCGCGTGGAGGACCTGTGATGACCGGAGAGCCTGACGCG 3289
QY 99 aGluAspArgArg 103
Db 3290 CGAGGACCGGCGG 3302

RESULT 3
US-10-126-927-67
/ Sequence 67, Application US/10126927
/ Publication No. US20030082575A1
/ GENERAL INFORMATION:
/ APPLICANT: The Scripps Research Institute
/ APPLICANT: Schultz, Peter G
/ APPLICANT: Wang, Lei
/ APPLICANT: Anderson, John C
/ APPLICANT: Chin, Jason
/ APPLICANT: Liu, David R
/ APPLICANT: Magliery, Thomas
/ APPLICANT: Meggers, Eric L
/ APPLICANT: Mehl, Ryan A
/ APPLICANT: Pasternak, Miro
/ APPLICANT: Santoro, Stephen W
/ APPLICANT: Zhang, Zhiwen
/ TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
/ FILE REFERENCE: 54-000120US
/ CURRENT APPLICATION NUMBER: US/10/126,927
/ CURRENT FILING DATE: 2002-04-19
/ PRIOR APPLICATION NUMBER: US 60/285,030
/ PRIOR FILING DATE: 2001-04-19
/ PRIOR APPLICATION NUMBER: US 60/355,514
/ PRIOR FILING DATE: 2002-02-06
/ NUMBER OF SEQ ID NOS: 79
/ SOFTWARE: Patentin version 3.1
/ SEQ ID NO 67
/ LENGTH: 12391
/ TYPE: DNA
/ ORGANISM: Plasmid pSci101, Streptomyces venezuelae papABC
US-10-126-927-67
```



Thu Oct 6 09:19:07 2005

us-10-089-514-4.rmpb

Page 3

## Alignment Scores:

Pred. No.: 6,28e-55 Length: 12391  
 Score: 439.00 Matches: 96  
 Percent Similarity: 94.12% Conservative: 0  
 Best Local Similarity: 94.12% Mismatches: 3  
 Query Match: 85.41% Indels: 3  
 DB: 14 Gaps: 2

US-10-089-514-4 (1-103) x US-10-126-927-67 (1-12391)

QY 1 MetThrGluGlnaAngluLeuGln---ArgLeuArgAlaGluLeuAspAlaLeuAspG1 19  
 DB 2055 ATGACCGACGACGAGAGCTGACAGTTCGCGCTCGCGAGAGCTCGACCCCTCGACGG 2114  
 QY 19 YThrLeuLeuAspThrAlaArgArgAlaLeuGlyValAlaArgIleAlaArgTyrIly 39  
 DB 2115 GACGCTTGTGACACGGTGGCGCGCCGATCGACTCGGTGTCCGATCCGCGGTACAA 2174  
 QY 39 sSerArgHisGlyValPrometMetGlnProGlyArgValSerLeuValIysAspArgAl 59  
 DB 2175 GTCCCGGACGCGCTCCGATGATGACGCCGCGGTCAAGCTGTCAAGACAGGCG 2234  
 QY 59 aAlaArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspVa 79  
 DB 2235 CGCCCGCTACGCGCGACACCGGCTCGACGAATCGTTCGTGAACCTCTACGACGT 2294  
 QY 79 I1IleIleThrGluMetCysArgValGluAspLeuValMetSerArgGluSerLeuThrAl 99  
 DB 2295 GATCATCATCGAGATGTGCGCGCTCGAGACCTGTGTATGAGC---CCGTCACTGTACTAA 2351  
 QY 99 aglu 100  
 DB 2352 GGAG 2355

## RESULT 4

US-10-137-310-1  
 ; Sequence 1, Application US/10137310  
 ; Publication No. US20030176675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Valentín, Henry E.  
 ; APPLICANT: Mitsky, Timothy A.  
 ; TITLE OF INVENTION: Tyra Genes and Uses Thereof  
 ; FILE REFERENCE: 16515.147  
 ; CURRENT APPLICATION NUMBER: US/10/137,310  
 ; CURRENT FILING DATE: 2002-05-03  
 ; PRIOR APPLICATION NUMBER: US 60/289,527  
 ; PRIOR FILING DATE: 2001-05-09  
 ; NUMBER OF SEQ ID NOS: 12  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 1  
 ; LENGTH: 1122  
 ; TYPE: DNA  
 ; ORGANISM: Erwinia herbicola  
 US-10-137-310-1

## Alignment Scores:

Pred. No.: 6.86e-05 Length: 1122  
 Score: 102.00 Matches: 27  
 Percent Similarity: 50.00% Conservative: 22  
 Best Local Similarity: 27.55% Mismatches: 33  
 Query Match: 19.84% Indels: 16  
 DB: 16 Gaps: 2

US-10-089-514-4 (1-103) x US-10-137-310-1 (1-1122)

QY 6 GluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrVal 25  
 DB 10 GAACGTACCGCGCTTACCGCATCAATTGACAGCTTAATAAGCGCTCGATCTGCTG 69  
 QY 26 ArgArgArgIleAspLeuGlyValArgIleAlaArgTyrIysSerArgHisGlyValPro 45  
 DB 70 GCTAAGCACTCGAAGTGTGGCCGAGTGTAGTGAAGACCGCTTACGCGCTGCT 129

QY 46 MetMetGlnProGlyArgValSerLeuValIysAspArgAlaAlaArgTyrAlaAlaAsp 65  
 DB 130 ATCTATGTGCTTACGCTGACGGCTGACAGTCTGCTTCCGCTGCAAAAGCGCAAGC 189  
 QY 66 HisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIleThrGluMetCys 85  
 DB 190 CTCGGGTACACCGGATCTG----- 210

QY 86 ArgValGluAspLeuVal-----MetSerArgGluSerLeuThrAlaGluAsp 101

DB 211 --ATTGAGATGTGCTGCTGCGTGATGCGGGAATCTATACGCGAGAA 261

## RESULT 5

US-10-634-548-16  
 ; Sequence 16, Application US/10634548  
 ; Publication No. US20040045051A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: No. US20040045051A1ris, Susan R  
 ; APPLICANT: Lincoln, Kim  
 ; APPLICANT: Abad, Mark Scott  
 ; APPLICANT: Eilers, Robert  
 ; APPLICANT: Hartsuiker, Karen Kindle  
 ; APPLICANT: Hirtsberg, Joseph  
 ; APPLICANT: Karunanadas, Balasulojini  
 ; APPLICANT: Moshiri, Farhad  
 ; APPLICANT: Stein, Joshua C.  
 ; APPLICANT: Valentín, Henry E.  
 ; APPLICANT: Venkatesh, Tyamagondlu V.  
 ; TITLE OF INVENTION: Tocopherol biosynthesis related genes and used thereof  
 ; FILE REFERENCE: Ren-01-125  
 ; CURRENT APPLICATION NUMBER: US/10/634,548  
 ; CURRENT FILING DATE: 2003-08-05  
 ; PRIOR APPLICATION NUMBER: US 60/400,689  
 ; PRIOR FILING DATE: 2002-08-05  
 ; NUMBER OF SEQ ID NOS: 79  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 16  
 ; LENGTH: 1122  
 ; TYPE: DNA  
 ; ORGANISM: Erwinia herbicola  
 US-10-634-548-16

## Alignment Scores:

Pred. No.: 6.86e-05 Length: 1122  
 Score: 102.00 Matches: 27  
 Percent Similarity: 50.00% Conservative: 22  
 Best Local Similarity: 27.55% Mismatches: 33  
 Query Match: 19.84% Indels: 16  
 DB: 18 Gaps: 2

US-10-089-514-4 (1-103) x US-10-634-548-16 (1-1122)

QY 6 GluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrVal 25  
 DB 10 GAACGTACCGCGCTTACCGCATCAATTGACAGCTTAATAAGCGCTCGATCTGCTG 69  
 QY 26 ArgArgArgIleAspLeuGlyValArgIleAlaArgTyrIysSerArgHisGlyValPro 45  
 DB 70 GCTAAGCACTCGAAGTGTGGCCGAGTGTGAGTGAAGCGCGTATACGCGCTGCT 129  
 QY 46 MetMetGlnProGlyArgValSerLeuValIysAspArgAlaAlaArgTyrAlaAlaAsp 65  
 DB 130 ATCTATGTGCTTACGCTGACGGCTGACAGTCTGCTTCCGCTGCAAAAGCGCAAGC 189  
 QY 66 HisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIleThrGluMetCys 85  
 DB 190 CTCGGGTACACCGGATCTG----- 210  
 QY 86 ArgValGluAspLeuVal-----MetSerArgGluSerLeuThrAlaGluAsp 101  
 DB 211 --ATTGAGATGTGCTGCTGCGTGATGCGGGAATCTTATACGCGAGAA 261

## RESULT 6







```
; SEQ ID NO 40
; LENGTH: 1833
; TYPE: DNA
; ORGANISM: Sphingomonas trueperi
US-10-381-779-40

Alignment Scores:
Pred. No.: 0.000943 Length: 1833
Score: 96.50 Matches: 27
Percent Similarity: 43.02% Conservative: 10
Best Local Similarity: 31.40% Mismatches: 48
Query Match: 18.77% Indels: 1
DB: 17 Gaps: 1

US-10-089-514-4 (1-103) x US-10-381-779-40 (1-1833)

QY 1 MetThrGlnGlnAsnGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20
   |||||:::|||||:::|||||:::|||||:::|||||:::
DB 522 ATGACCGACGCGACG--CTGACCCCGCTATCGCCAGACATCGACATATCGATGCGCGC 466

QY 21 LeuLeuAspThrValArgArgGlyIleAspLeuGlyValArgIleAlaArgTyrLysSer 40
   |||||:::|||||:::|||||:::|||||:::|||||:::
DB 465 CTGCTCTTCATGCTCGCCGACGGTTCCAAAGTGAACCGACGGCGGTGGCGCTTCAAGGCC 406

QY 41 ArgHisGlyValProMetMetGlnProGlyArgValSerLeuValLysAspArgAlaAla 60
   |||||:::|||||:::|||||:::|||||:::|||||:::
DB 405 GAATCCGGGCTGCCCGCCGCGATCCGCGCCGACGAGGCGCAGATCGCCCGGCTGCCG 346

QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIle 80
   |||||:::|||||:::|||||:::|||||:::|||||:::
DB 345 GCGCTGGCGCAGGATGCGCATCTCGACCCGAATTTCCGAGAATTCCTGCGCTTCATC 286

QY 81 IleThrGluMetCysArg 86
   |||||:::|||||:::|||||:::|||||:::
DB 285 ATCAGCAGGTGATCCGC 268

RESULT 9
US-10-156-761-6187
; Sequence 6187, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6187
; LENGTH: 387
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(387)
US-10-156-761-6187

Alignment Scores:
Pred. No.: 0.000742 Length: 387
Score: 91.00 Matches: 24
Percent Similarity: 45.68% Conservative: 13
Best Local Similarity: 29.63% Mismatches: 44
Query Match: 17.70% Indels: 0
DB: 15 Gaps: 0
```

```
US-10-089-514-4 (1-103) x US-10-156-761-6187 (1-387)

QY 6 GluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrVal 25
   |||||:::|||||:::|||||:::|||||:::|||||:::
DB 7459246 GAATCGCCCGGCTGCCGACGATCGACATCGACGCGCGCTCTCCACATGCTC 7459305

QY 26 ArgArgArgIleAspLeuGlyValArgIleAlaArgTyrLysSerArgHisGlyValPro 45
   |||||:::|||||:::|||||:::|||||:::|||||:::
DB 118 GCCGAAGCTTCAAGTGCACCCAGCAGTCCGACCTCAAGCCGTACACAGCTGGCG 177

QY 46 MetMetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyrAlaAlaAsp 65
   |||||:::|||||:::|||||:::|||||:::|||||:::
DB 178 CCGCGCGACCGCGCGCGCGCAGGCGCAGATCGCCGACGCTGCGCGCGCGAGAAC 237

QY 66 HisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIleThrGluMetCys 85
   |||||:::|||||:::|||||:::|||||:::|||||:::
DB 238 GCCAACTCGACCGCGCGCTTCGCGGAGAGCTGCTGGCTTCATCTCCGCGAGTTCATC 297

QY 86 Arg 86
   |||||:::|||||:::|||||:::|||||:::
DB 298 CGC 300

RESULT 10
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 955 Length: 9025608
Score: 91.00 Matches: 24
Percent Similarity: 45.68% Conservative: 13
Best Local Similarity: 29.63% Mismatches: 44
Query Match: 17.70% Indels: 0
DB: 15 Gaps: 0

US-10-089-514-4 (1-103) x US-10-156-761-1 (1-9025608)

QY 6 GluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrVal 25
   |||||:::|||||:::|||||:::|||||:::|||||:::
DB 7459246 GAATCGCCCGGCTGCCGACGATCGACATCGACGCGCGCTCTCCACATGCTC 7459305

QY 26 ArgArgArgIleAspLeuGlyValArgIleAlaArgTyrLysSerArgHisGlyValPro 45
   |||||:::|||||:::|||||:::|||||:::|||||:::
DB 7459306 GCCGAAGCTTCAAGTGCACCCAGCAGTCCGCGCACCTCAAGCCGTACACAGCTGGCG 7459365

QY 46 MetMetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyrAlaAlaAsp 65
   |||||:::|||||:::|||||:::|||||:::|||||:::
DB 178 CCGCGCGACCGCGCGCGCGCAGGCGCAGATCGCCGACGCTGCGCGCGAGAAC 237
```



```

Db      7459366 CCCGCCGACCCGCCGCGAGCCAGATCGCCGCTGCGCGCTCGCGAGAAC 7459425
Qy      66 HieglYleuaspJuserPheleuValAsnleuTyraValIleIleThrgIuMetCys 85
Db      7459426 GCCCAACTCGACCCGGCGTTTCGCCGAGAACTGCTGGGCTTCATCATCGCCGAGTCAATC 7459485
Qy      86 Arg 86
Db      7459486 CGC 7459488

RESULT 11
US-10-137-310-3
; Sequence 3, Application US/10137310
; Publication No. US20030176675A1
; GENERAL INFORMATION:
; APPLICANT: Valentin, Henry E.
; APPLICANT: Mitsky, Timothy A.
; TITLE OF INVENTION: Tyra Genes and Uses Thereof
; FILE REFERENCE: 16515.147
; CURRENT APPLICATION NUMBER: US/10/137,310
; PRIOR FILING DATE: 2002-05-03
; PRIOR APPLICATION NUMBER: US 60/289,527
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1122
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-137-310-3

Alignment Scores:
Pred. No.:      0.00467      Length:      1122
Score:          90.00      Matches:      28
Percent Similarity: 49.49%      Conservative: 21
Best Local Similarity: 28.28%      Mismatches: 32
Query Match:    17.51%      Indels:      18
                        Gaps:      4

US-10-089-514-4 (1-103) x US-10-137-310-3 (1-1122)

Qy      6 GluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrVal 25
Db      10 GAATTGACCCGATACCGCATCAAAATTGATGAAGTCATAACCGCTGCGATTATTA 69
Qy      26 ArgArgArgIleAspLeuGlyValArgIleAlaArgTyrLysSerArgHisGlyValPro 45
Db      70 GCGAAGCGTCTGGAACCTGTTGCTGAAGTGGCGAGGTGAAAGCCGCTTGACTGCTCT 129
Qy      46 MetMetGlnProGlyArg--ValSerLeuValLysAspArgAlaAlaArgTyrAlaAla 64
Db      130 ATTATGTTCCGAGAGCGAGAGCATCTATGTTGGCCCTCGCTGTCGAGAG--GCGGAA 186
Qy      65 AspHisGlyLeuAspJuserPheleuValAsnleuTyraValIleIleThrgIuMet 84
Db      187 GCTCTGGGTACCGCAGATCTG-----MetSerArgGluSerLeuThrAlaGluAsp 210
Qy      85 CysArgValGluSerLeuVal-----MetSerArgGluSerLeuThrAlaGluAsp 101
Db      211 -----ATTGAGAGATGTTTGGCTCGGGTGATGCGTGAATCTTACTCCAGTGAAC 261

RESULT 12
US-10-433-556-18
; Sequence 18, Application US/10433556
; Publication No. US20040117872A1
; GENERAL INFORMATION:
; APPLICANT: Matringe, Michel
; APPLICANT: Rippert, Pascal
; TITLE OF INVENTION: Novel Targets For Herbicides And Transgenic Plants Resistant To
; TITLE OF INVENTION: Said Herbicides
; FILE REFERENCE: 5500-120
; CURRENT APPLICATION NUMBER: US/10/433,556
```

```

; CURRENT FILING DATE: 2003-06-04
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
; LENGTH: 1122
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1122)
US-10-433-556-18

Alignment Scores:
Pred. No.:      0.00467      Length:      1122
Score:          90.00      Matches:      28
Percent Similarity: 49.49%      Conservative: 21
Best Local Similarity: 28.28%      Mismatches: 32
Query Match:    17.51%      Indels:      18
                        Gaps:      4

US-10-089-514-4 (1-103) x US-10-433-556-18 (1-1122)

Qy      6 GluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrVal 25
Db      10 GAATTGACCCGATACCGCATCAAAATTGATGAAGTCATAACCGCTGCGATTATTA 69
Qy      26 ArgArgArgIleAspLeuGlyValArgIleAlaArgTyrLysSerArgHisGlyValPro 45
Db      70 GCGAAGCGTCTGGAACCTGTTGCTGAAGTGGCGAGGTGAAAGCCGCTTGACTGCTCT 129
Qy      46 MetMetGlnProGlyArg--ValSerLeuValLysAspArgAlaAlaArgTyrAlaAla 64
Db      130 ATTATGTTCCGAGAGCGAGAGCATCTATGTTGGCCCTCGCTGTCGAGAG--GCGGAA 186
Qy      65 AspHisGlyLeuAspJuserPheleuValAsnleuTyraValIleIleThrgIuMet 84
Db      187 GCTCTGGGTACCGCAGATCTG-----MetSerArgGluSerLeuThrAlaGluAsp 210
Qy      85 CysArgValGluSerLeuVal-----MetSerArgGluSerLeuThrAlaGluAsp 101
Db      211 -----ATTGAGAGATGTTTGGCTCGGGTGATGCGTGAATCTTACTCCAGTGAAC 261

RESULT 13
US-10-380-132-7
; Sequence 7, Application US/10380132
; Publication No. US20030182679A1
; GENERAL INFORMATION:
; APPLICANT: SunGene GmbH & Co. KGaA
; TITLE OF INVENTION: Improved processes for vitamin E biosynthesis
; FILE REFERENCE: NAE445/2000
; CURRENT APPLICATION NUMBER: US/10/380,132
; CURRENT FILING DATE: 2003-03-11
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1238
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: gene
; LOCATION: (7)..(1232)
; OTHER INFORMATION: Tyra gene coding for bifunctional chorismate
; OTHER INFORMATION: mutase / prephenate dehydrogenase
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (25)..(1143)
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(6)
; OTHER INFORMATION: restriction site linker
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1233)..(1238)
```



```

; OTHER INFORMATION: restriction site linker
US-10-380-132-7

Alignment Scores:
Pred. No.: 0.00536      Length: 1238
Score: 90.00           Matches: 28
Percent Similarity: 49.49%      Conservative: 21
Best Local Similarity: 28.28%    Mismatches: 32
Query Match: 17.51%           Indels: 18
DB: 16                  Gaps: 4

US-10-089-514-4 (1-103) x US-10-380-132-7 (1-1238)

QY 6 GLULEUQINARGLUARGALAGLULEUASPALAEUASPOLYTHRIEUEUASPTHRVAL 25
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 34 GAATTGACCCGATTCACCGATCAATGATGAAGTCAATGAAGCGCTGCTGAATTATT 93
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 26 ARGARGRGLLEASPLEUGLYVALARGILEALARGTYRYSERARGHISGLYVALPRO 45
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 94 GCGAAGCGTCTGGAACCTGTTGCTGAAGTGGCGAGGTGAAGCCGCTTGACCTGCT 153
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 46 METMEGLNPROGLIARG--VALSERLEUVALLYSAPARGALALARGTYRALALA 64
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 154 ATTATATCTCCGAGCGCGAGCATCTATCTTGGCTCGCGCTCGAGAG--GCCGAA 210
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 65 ASPHISGLYEUASPOLYSERPHELEUVALASPLEUTYRASPVALLIETHRGLUMET 84
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 211 GCTCTGGGTGTACCCGACATCTG----- 234
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 85 CYASRGVALGLUASPLEUVAL-----METSERARGLSERLEUTHRALAGLUASP 101
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 235 -----ATTGAGATGTTTGGCGTGGGTGATGCTGAATCTTACTCAGTGAAGAAC 285
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 14
US-10-282-122A-15346
; Sequence 15346, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreysht, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308

```

```

; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed : See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15346
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-15346

Alignment Scores:
Pred. No.: 0.00545      Length: 1104
Score: 89.50           Matches: 19
Percent Similarity: 50.65%      Conservative: 20
Best Local Similarity: 24.68%    Mismatches: 35
Query Match: 17.41%           Indels: 3
DB: 17                  Gaps: 1

US-10-089-514-4 (1-103) x US-10-282-122A-15346 (1-1104)

QY 6 GLULEUQINARGLUARGALAGLULEUASPALAEUASPOLYTHRIEUEUASPTHRVAL 25
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 7 GAACCTGAAGAAATAAGAAAGAAATAAGACGAGTGAATAAAATTATACAGAGCTTTT 66
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 26 ARGARGRGLLEASPLEUGLYVALARGILEALARGTYRYSERARGHISGLYVALPRO 45
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 67 GAAAAAGATGATTAATTTGTTATGAGATGCTGAGTATTAAGAAAAAATTAGTATTTCCA 126
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 46 METMEGLNPROGLIARGVALSERLEUVALLYSAPARGALALARGTYRALALAASP 65
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 127 GTTAGTGATGACAAAGAGAAAAAAGTTATGATATAAAATATAGTTACTTAAGAAAT 186
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 66 HISGLYEUASPOLY-----SERPHELEUVALASPLEUTYRASPVALLIETHRGLUMET 79
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 187 AAGCAATATCTGAGGTTCTTAAGAGATTTATATATTAATTTATGAACCTTA 237
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 15
US-10-329-670-1
; Sequence 1, Application US/10329670
; Publication No. US20040018503A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleic Acid Sequence of the Haemophilus influenzae Rd Genome, Fra
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,670
; CURRENT FILING DATE: 2002-12-24
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature

```



```
/ LOCATION: (29298)..(29298)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (36543)..(36543)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (36551)..(36551)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (36636)..(36636)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (40808)..(40810)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (44416)..(44416)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (44905)..(44905)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (44975)..(44975)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (45593)..(45593)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (45732)..(45732)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (47036)..(47036)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51334)..(51334)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51602)..(51602)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51786)..(51786)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51805)..(51805)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (55369)..(55369)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (65309)..(65309)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (65313)..(65313)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (80024)..(80024)

/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (10091)..(10091)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (102696)..(102696)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (105121)..(105121)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (107248)..(107248)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (117136)..(117136)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (119750)..(119750)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (119924)..(119924)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (120038)..(120038)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (121344)..(121344)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (122167)..(122167)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (122316)..(122316)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (131340)..(131340)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (131360)..(131360)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (139910)..(139910)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (140398)..(140398)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (142750)..(142750)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (145058)..(145058)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (145171)..(145171)
/ OTHER INFORMATION: n equals a, t, g or c
```



```

FEATURE:
NAME/KEY: misc feature
LOCATION: (145942) .. (145942)
OTHER_INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (147197) .. (147197)
OTHER_INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (150841) .. (150841)
OTHER_INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152500) .. (152500)
OTHER_INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152530) .. (152530)

```

Pred. No.:	183012
Score:	2.57
Length:	88.50
Percent Similarity:	67.92%
Percent Similarity:	33.08%
Best Local Similarity:	17.22%
Query Match:	17
DB:	1
Length:	183012
Matches:	17
Conservative:	19
Mismatches:	16
Indels:	1
Gaps:	1

QY	Db	QY	Db	QY	Db
7	1370466	27	1370526	47	1370586
LeuGlnArgLeuArgIleGluIleuAspAlaLeuAspGlyThrLeuLeuAspThrValArg	CTTAAAGATTACGTTCTGTGAATATGATTCGCTTATCGCGAACTTATCAACTTTTGCT	ArgArgIleAspLeuGluValArgIleAlaArgTyrTrpSerAspGlnIleValPromet	AAACGCTTGAAGTTGCTTTCTCAAGTCGCGTAAAGTGAACACTCAATGATTTACTATT	MetGlnProGlyArg--ValSerLeuValIleAspArg	TATGCGCGAAGCGTGAATTAGCAATCTCTCAACACAGT
26	1370526	46	1370586	58	1370624

Search completed: October 6, 2005, 01:50:33  
Job time : 5167.67 secs



This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 09:03:04 ; Search time 1126.79 Seconds  
(without alignments)  
3479.464 Million cell updates/sec

Title: US-10-089-514-4

Perfect score: 514

Sequence: 1 MTEQNELQRLRAELDALDGT.....MCRVEDLVMSRESLTAEDDR 103

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues  
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool.h/US10089514/runat\_04102005\_105744\_8043/app\_query.fasta\_1.1621  
-DB=EST -QFMT=fastcap -SUFFIX=rcst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS-bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pct -NOR=ext -HEA8R125=500 -MINLEN=2000000000  
-USR=US10089514.@CGN\_1\_1\_6050.@runat\_04102005\_105744\_8043 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
EST:.\*  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hic.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_gss1.\*  
9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85.5	16.6	806	9	CL688507 PRI0149C
2	82	16.0	846	7	CN822873 Oa.sp1bn
3	79.5	15.5	600	8	BZ895453 NaRpl1.01
4	79	15.4	471	2	AM637288 b156h07.w
5	78	15.2	431	9	CL674789 PRI0113a
6	77.5	15.1	509	4	BM326074 PIC1_65_E
7	77	15.0	692	8	BM454175 BOCI86TF
8	76.5	14.9	913	5	BU915943 AGENCOURT
9	75.5	14.7	627	7	CN894351 010625MAY

C	10	75	14.6	529	4	BM595564	BM595564 17006874
C	11	75	14.6	644	8	BH654655	BH654655 BOHTE10TF
C	12	75	14.6	756	8	BH248862	BH248862 BOGAM09TF
C	13	74	14.4	312	6	CD602770	CD602770 RK354A2F1
C	14	74	14.4	325	6	CD602824	CD602824 RK354A4G1
C	15	74	14.4	535	7	CF857652	CF857652 p8MY0021F
C	16	74	14.4	536	2	BE583221	BE583221 2-3F-MY P
C	17	74	14.4	559	5	BX601867	BX601867 BX601867
C	18	74	14.4	615	9	CL708793	CL708793 OR_BBA003
C	19	74	14.4	739	9	AL860917	AL860917 w156D07.x
C	20	74	14.4	775	8	BZ577997	BZ577997 m8n2_5671
C	21	74	14.4	775	9	CL727759	CL727759 OR_BBA005
C	22	73.5	14.3	368	6	CD061523	CD061523 MA1-0020P
C	23	73.5	14.3	569	1	AA941215	AA941215 LD5234.5
C	24	73.5	14.3	595	7	CF429729	CF429729 PH1_24_H0
C	25	73.5	14.3	624	7	CV233416	CV233416 WS01210.B
C	26	73.5	14.3	633	7	CV238063	CV238063 WS0125.B2
C	27	73.5	14.3	662	7	CV237417	CV237417 WS01227.B
C	28	73.5	14.3	685	7	CV237350	CV237350 WS01227.B
C	29	73.5	14.3	691	7	CV236915	CV236915 WS01225.B
C	30	73.5	14.3	727	9	CG840622	CG840622 Ynhw1236
C	31	73.5	14.3	909	4	BG343273	BG343273 HYSME5000
C	32	73	14.2	657	8	BH934632	BH934632 odt87C04.
C	33	73	14.2	693	8	BH576490	BH576490 BOGGA03TF
C	34	73	14.2	769	8	BH250514	BH250514 BOGAU29TF
C	35	73	14.2	789	8	BH436424	BH436424 BOHJ044TF
C	36	73	14.2	908	7	CF700884	CF700884 CCNF289TO
C	37	73	14.2	1019	7	CF706740	CF706740 CAC240TO
C	38	72.5	14.1	460	7	CO312074	CO312074 EK227832.
C	39	72.5	14.1	505	1	AA804082	AA804082 GM12763.5
C	40	72.5	14.1	530	2	BF494794	BF494794 AT03534.5
C	41	72.5	14.1	562	1	AL113596	AL113596 GH09960.5
C	42	72.5	14.1	572	4	B1355453	B1355453 CM30880.5
C	43	72.5	14.1	674	1	AI106885	AI106885 GH06135.5
C	44	72.5	14.1	712	2	BF493712	BF493712 AT02044.5
C	45	72.5	14.1	738	9	CNS06UH2	AL415869 T3 end of

## ALIGNMENTS

RESULT 1  
CL688507  
LOCUS  
DEFINITION  
CL688507 806 bp DNA linear GSS 09-JUN-2004  
PRI0149C.F10.2 - PRI0149C.BR (806) Mixed stage fosmid library of P.  
pacificus var. California Pristionchus pacificus genomic, genomic  
survey sequence.

ACCESSION  
CL688507  
VERSION  
CL688507.1 GI:50197585  
KEYWORDS  
GSS.

SOURCE  
Pristionchus pacificus  
ORGANISM  
Pristionchus pacificus

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 806)  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 806)  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 806)  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 806)  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 806)  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 806)  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 806)  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 806)  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 806)  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 806)  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 806)  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 806)  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 806)  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 806)  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 806)  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 806)  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 806)  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 806)  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 806)  
AparDB: an AcceB database for the nematode satellite organism

JOURNAL  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

REFERENCE  
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.



ORIGIN  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Mixed stage fosmid library of P. pacificus  
 var. California"  
 /note="Vector: pBplfos-5 Fosmid vector"

Alignment Scores:  
 Pred. No.: 0.608 Length: 806  
 Score: 85.50 Matches: 19  
 Percent Similarity: 59.26% Conservative: 13  
 Best Local Similarity: 35.19% Mismatches: 21  
 Query Match: 16.63% Indels: 1  
 DB: 9 Gaps: 1

US-10-089-514-4 (1-103) x CL688507 (1-806)

QY 6 GlnLeuGlnArgLeuAlaGlnLeuAspAlaLeuAspGlyThrLeuLeuAspThrVal 25

Db 640 GAATTCACCGCATTCACCGCATCAATTGATGAAGTCAAGCGCTGGAATTATTA 699

QY 26 ArgArgArgIleAspLeuGlyValArgIleAlaArgTyrIleSerArgHisGlyValPro 45

Db 700 GCGAAGCGTCTGGAACTGTTGCTGAAGTGGCGAGCGTGAAGACCGCTTTGGACGTGCT 759

QY 46 MetMetClnProGlyArg--ValSerLeuValLysAspArg 58

Db 760 ATTATGTTCCGAGCGCGAGCATCTATGTTGGCTCGCGG 801

RESULT 2

CN822873/c 846 bp mRNA linear EST 02-JUN-2004

LOCUS CN822873  
 DEFINITION Oa脾ln\_02E15\_M13reverse Sheep spleen\brain pSport1 library Ovis  
 aries cDNA clone Oa脾ln\_02E15 5', mRNA sequence.

ACCESSION CN822873

VERSION CN822873.1

KEYWORDS GI:47950942

SOURCE EST.

ORGANISM Ovis aries (sheep)

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Caprinae; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Caprinae; Ovis.

AUTHORS 1 (bases 1 to 846)

TITLE Gossner, A. and Hopkins, J.

JOURNAL Ovine spleen\brain cDNA library

COMMENT Unpublished (2004)

Contact: J Hopkins

Veterinary Biomedical Sciences

University of Edinburgh

Summerhall Square, Edinburgh, EH9 1QH.

Email: j.hopkins@ed.ac.uk

Plate: 02 row: E column: 15

Seq primer: M13reverse

High quality sequence start: 4

High quality sequence stop: 549.

Location/Qualifiers

1. .846

/organism="Ovis aries"

/mol\_type="mRNA"

/db\_xref="taxon:9940"

/clone="Oa脾ln\_02E15"

/clone\_lib="Sheep spleen\brain pSport1 library"

/note="Vector: pSport1"

ORIGIN

Alignment Scores:

Pred. No.: 1.94 Length: 846

Score: 82.00 Matches: 32

Percent Similarity: 46.30% Conservative: 18

Best Local Similarity: 29.63% Mismatches: 28

Query Match: 15.95% Indels: 30

DB: 7 Gaps: 7

US-10-089-514-4 (1-103) x CN822873 (1-846)

QY 4 GlnAspGlnLeuGlnArgLeuArgAlaGlnLeuAspAlaLeuAspGlyThrLeu----- 21

Db 689 CACCGCATCAGCAGCGCCACACAGCGCCACCGCCACCA---CATGGGGAAATTCACATC 633

QY 22 -----LeuAspThrValArgArgIleAspLeuGlyVal-----ArgIleAlaArg 37

Db 632 GTGCATCTTCGCGTCCGACGAGCGCGGTGATCTTCATGTACTGTCAGCGTGTGGCAC 573

QY 38 TyrIleSerArgHisGlyVal-----PrometMet----- 47

Db 572 CATCTGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTC 513

QY 48 -----GlnProGlyValSerLeuValLysAspArgAlaAlaArgTyrAlaAla 64

Db 512 GTTAGCGCAACATCCAGAGCGCGTTCGCGTTC-----CACGATGCGC 474

QY 65 AspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIleThrGluMet 84

Db 473 GACCATCGCGTTCGATCAGGCGCTTC-----CTTTGGGACCGTGTCCAGGAACTT 423

QY 85 CysArgValGlnAspLeuValMet 92

Db 422 CATCTGCTCGAACAGCTGTTCGTA 399

RESULT 3

BZ895453 600 bp DNA linear GSS 30-JUL-2003

LOCUS NARP11\_0120 Na pUC18 library Natrialba asiatica genomic 5', genomic

survey sequence.

ACCESSION BZ895453

VERSION BZ895453.1

KEYWORDS GI:33345929

SOURCE GSS.

ORGANISM Natrialba asiatica

REFERENCE Natrialba asiatica

AUTHORS Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 Halobacteriaceae; Natrialba.

TITLE 1 (bases 1 to 600)

JOURNAL Go, Y., Roach, J., Glusman, G., Baliga, N.S., Deutsch, K., Pan, M.,  
 Dasgupta, S., Ng, M.V. and Hood, L.

COMMENT Low-pass Sequencing for Microbial Comparative Genomics  
 Unpublished (2003)

Contact: Go, Y

Institute for Systems Biology

1441 North 34th Street, Seattle, WA 98103, USA

Tel: 206 732 1412

Fax: 206 732 1299

Email: ygoo@systemsbiology.org

Seq primer: M13 Forward

Class: shotgun.

Location/Qualifiers

1..600

/organism="Natrialba asiatica"

/mol\_type="genomic DNA"

/strain="ATCC 700177"

/db\_xref="taxon:64602"

/clone\_lib="Na pUC18 library"

/note="Vector: pUC18, Site\_1: SmaI; A shotgun library was  
 constructed from Natrialba asiatica genomic DNA using  
 pC18/SmaI/BAP plasmid"

ORIGIN

Alignment Scores:

Pred. No.: 2.78 Length: 600

Score: 79.50 Matches: 22

Percent Similarity: 51.06% Conservative: 26

Best Local Similarity: 23.40% Mismatches: 41

Query Match: 15.47% Indels: 5

DB: 8 Gaps: 2

US-10-089-514-4 (1-103) x BZ895453 (1-600)



QY 3 GluGlnAnsgluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeu 22  
Db 28 GAGCAGATGATCTCGAACAACCTCCGAGAGATGATGACATCATCGAGATGCTC 87  
QY 23 AspThrValArgArgArgGlyLeuAspGlyValArgIleAlaArgIlyrLySerArgHis 42  
Db 88 GAACGTGATCCGCGCAGCACTACGTGCGGATTCGATTCGACGGTCAAGCAGCAGAG 147  
QY 43 G1ValPrometMetGlnProG1ArgValSerLeuValysAspArgAlaAlaArgTyr 62  
Db 148 GACCTGCCGACGACCGCAGACACCATGATGACAGTAAATGCAACGACGCGGCGCAGAAC 207  
QY 63 AlaAlaAspHisg1LeuAspGluSerPheLeuValAsnLeuTyrAspValIleIleThr 82  
Db 208 GCAGAGACGATTGCGATCGACGCGAATCTGTGTGAAGCGCATCTTCGTTATTGATC--- 264  
QY 83 GluMetCysArgValGluAspLeuValMetSerArgGluSer 96  
Db 265 GAACGTGAACAAGTTGAG-----CAGCGCAGTCC 294  
RESULT 4  
AM637288 471 bp mRNA linear EST 26-APR-2001  
LOCUS b156h07.w1 Blackshear/Soares normalized Xenopus egg library Xenopus  
DEFINITION laevis cDNA clone PBX0056H07 5', mRNA sequence.  
ACCESSION AM637288  
VERSION AM637288.1 GI:7394396  
KEYWORDS EST.  
SOURCE Xenopus laevis (African clawed frog)  
ORGANISM Xenopus laevis  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
Xenopodidae; Xenopus; Xenopus.  
1 (bases 1 to 471)  
Blackshear, P.J., Lai, W.S., Thorn, J.M., Kennington, E.A., Staffa, N.G.  
Jr., Moore, D.T., Bouffard, G.G., Beckstrom-Sternberg, S.M.,  
Touchman, J.W., Bonaldo, M.F. and Soares, M.B.  
The NIHES Xenopus maternal EST project: interim analysis of the  
first 13,879 ESTs from unfertilized eggs  
Gene 267 (1), 71-87 (2001)  
21311403  
11311557  
JOURNAL Contact: Perry J. Blackshear  
MEDLINE Office of Clinical Research and Laboratory of Signal Transduction  
PUBMED National Institute of Environmental Health Sciences  
COMMENT A2-05 NIHES, 101 Alexander Drive, Research Triangle Park, NC 27709,  
USA  
Tel: 919 541-4899  
Fax: 919 541-4571  
Email: black009@niehs.nih.gov  
Clone is available through Research Genetics, Inc., 2130 Memorial  
Parkway, Huntsville, AL 35901  
phone 800-533-4363 ext.cdna, fax 256-536-9016 alt:cdna, email  
cna@resgen.com  
DNA Sequencing and analyses performed by National Institutes of  
Health Intramural Sequencing Center (NISC).  
PCR Primers  
FORWARD: TGTAAACGACGCGCCAGT  
BACKWARD: CAGAAACAGCTATGACC  
Plate: 0056 row: H column: 07  
Seq primer: T7 primer.  
Location/Qualifiers  
1..471  
/organism="Xenopus laevis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8355"  
/clone="PBX0056H07"  
/sex="female"  
/tissue\_type="unfertilized egg"  
/cell\_type="unfertilized egg"  
/dev\_stage="unfertilized egg"  
/lab\_host="DH108"  
/clone\_lib="Blackshear/Soares normalized Xenopus egg"

library"  
/note="Vector: pT73-Pac; Site 1: EcoRI; Site 2: NotI;  
polyA-selected mRNA was prepared from unfertilized Xenopus  
laevis eggs. The library was constructed in the vector  
pT73-Pac as described in Bonaldo, M.F., Jenson, G. and  
Soares, M.B. 'Normalization and subtraction: two  
approaches to facilitate gene discovery', Genome Research  
6:791-806, 1996. The first strand synthesis used a  
NotI-dt18 primer; double stranded cDNAs were ligated to  
EcoRI adapters, digested with NotI, and directionally  
cloned into the NotI and EcoRI-digested pT73-Pac vector.  
The library contained approximately 7.2 x 10<sup>5</sup>  
recombinants, with average insert sizes of 1-1.5 kb."

ORIGIN  
Alignment Scores:  
Pred. No.: 2.4 Length: 471  
Score: 79.00 Matches: 28  
Percent Similarity: 50.48 Conservative: 25  
Best Local Similarity: 26.67 Mismatches: 38  
Query Match: 15.37 Indels: 14  
DB: 2 Gaps: 4  
US-10-089-514-4 (1-103) x AM637288 (1-471)  
QY 4 GlnAnsgluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr----- 20  
Db 10 AGGCTGAAATTAATAAGCTGAAGGCCACGCTTCACAGCGTGAAGTCAACCAAGAC 69  
QY 21 LeuLeuAspThrValArgArgArgIle-----AspLeuGlyValArgIle 35  
Db 70 GTCCTGATGGATGATGAGAAAGATCTCACAGCTCACTCAGAGTTCACCGAGAGAT 129  
QY 36 AlaArgTyrLySerArgHisg1ValPrometMetGlnProG1ArgValSerLeuVal 55  
Db 130 ATCAGATGCGCTTCACGTCAGCATGACCTGCGAAGG-----AAATTAAGCCGAA 183  
QY 56 LysAspArgAlaAlaArgTyrAlaAlaAspHisg1LeuAspGluSerPheLeuValAsn 75  
Db 184 CGGCAGAGGCGCAGAGAGAGGCTGCAGAACAGAGATGCTTAAATGCTCTGGAAGAAC 243  
QY 76 LeuTyrAspValIleIleThrGluMetCysArgValGluAspLeuValMetSerArgGlu 95  
Db 244 TTGAGGACG-----GAGATTGCCGGTTATCGGAATCTTCAGACGCAAGAG 291  
QY 96 SerLeuThrAlaGlu 100  
Db 292 CCGGATGTGCGCAG 306  
RESULT 5  
CL674789 431 bp DNA linear GSS 09-JUL-2004  
LOCUS PRI0113a.E03.2 - PRI0113a.BR (431) Mixed stage fosmid library of P.  
DEFINITION pacificus var. California Pristionchus pacificus genomic  
survey sequence.  
ACCESSION CL674789  
VERSION CL674789.1 GI:50178309  
KEYWORDS Pristionchus pacificus  
SOURCE Pristionchus pacificus  
ORGANISM Pristionchus pacificus  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus:  
1 (bases 1 to 431)  
Srinivasan, J., Otto, G.W., Kahlow, U., Geisler, R. and Sommer, R.J.  
ApADB: an Acedb database for the nematode satellite organism  
Pristionchus pacificus  
Nucleic Acids Res. 32 (1), D421-D422 (2004)  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498



Email: ralf.sommer@tuebingen.mpg.de  
This library was generated at Caltech, Pasadena, USA and end  
sequenced at Vancouver, Canada.  
Seq primer: T7

Class: fosmid ends.

FEATURES  
source  
Location/Qualifiers  
1..431

/organism="Pristionchus pacificus"  
/mol\_type="genomic DNA"  
/strain="California"  
/db\_xref="taxon:54126"  
/clone\_lib="Mixed stage fosmid library of P. pacificus  
var. California"  
/note="Vector: pepifos-5 Fosmid vector"

## ORIGIN

## Alignment Scores:

Pred. No.: 2.95 Length: 431  
Score: 78.00 Matches: 23  
Percent Similarity: 45.68% Conservative: 14  
Best Local Similarity: 28.40% Mismatches: 44  
Query Match: 15.18% Indels: 0  
DB: 9 Gaps: 0

US-10-089-514-4 (1-103) x CL674789 (1-431)

QY 1 MetThGlnGlnAsnGluGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20  
Db 143 ATGACATCGGAATACCGCTTACTGGCGCTCGAGAGAAATCAGCGCGCTGATGAAAAA 202  
QY 21 LeuLeuAspThrValAlaArgArgAlaGlyLeuAspGlyValArgIleAlaArgTyrLysSer 40  
Db 203 TTATTAGCATTAATCTCCAGACGCGCGCACTGCGCTCGAGGAAAGCAAACTG 262  
QY 41 ArgHisGlyValProMetMetGlnProGlyArgValSerLeuValLysAspArgAlaAla 60  
Db 263 CTCCTCCGATCGCCCGGACGTGATTTTTCGAAACCCATTTACTCGAAAGATTAATT 322  
QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIle 80  
Db 323 ACGCTCGGTAAACGCGACCATCTGAGCGCCCATTAATTCTCGCGTTCGACGTCAATC 382  
QY 81 Ile 81  
Db 383 ATT 385

## RESULT 6

BM326074/c 509 bp mRNA linear EST 04-JAN-2002

LOCUS PIC1\_65\_E01.b1 A002 Pathogen-infected compatible 1 (PIC1) Sorghum

DEFINITION bicolor cDNA, mRNA sequence.

ACCESSION BM326074

KEYWORDS BM326074.1 GI:18065211

SOURCE EST.

ORGANISM Sorghum bicolor (sorghum)

Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoidae; Andropogoneae; Sorghum.

1 (bases 1 to 509)  
Cordonnier-Pratt, M.-M., Gingle, A., Fang, G.C., Dean, R., Wing, R.,  
Sudman, M., and Pratt, L.H.

An EST database from Sorghum: plants infected with a compatible  
pathogen

Unpublished (2002)

CONTACT: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmp@tuebingen.mpg.de  
Sequences have been trimmed to exclude PolyA, vector, and regions

below phred quality 16. The threshold for highest quality sequence  
is 20. Three-prime sequences, which are obtained with PolyTmX or  
T7 sequencing primer, are presented as the reverse complement.

Seq primer: JEN REV

High quality sequence stop: 433

PolyA=No.

FEATURES  
source  
Location/Qualifiers  
1..509

/organism="Sorghum bicolor"  
/mol\_type="mRNA"  
/cullivar="BTx623"  
/db\_xref="taxon:4558"  
/issue\_type="Leaves"  
/dev\_stage="4-week-old seedlings infected with  
Colletotrichum graminicola"  
/clone\_lib="Pathogen-infected compatible 1 (PIC1)"  
/note="Vector: pBluescript II SK(-) from lambda Zap II;  
Site\_1: XhoI; Site\_2: EcoRI; Four-week-old sorghum  
seedlings were sprayed with spore suspension prepared from  
3-week-old FRM421, a sorghum isolate of the anthracnose  
pathogen Colletotrichum graminicola. Inoculated plants  
were kept in a 25 C dark growth chamber with 100% relative  
humidity for 24 hr, followed by 12/12 hr of light/dark  
cycle at 25 C with 90% relative humidity for another 24  
hr. All leaves were harvested and quick frozen with liquid  
nitrogen and stored in a -80 C freezer. The library was  
made from poly-A RNA in the cloning vector lambda Zap II.  
Clones to be sequenced were prepared by mass excision.  
WARNING: While most or all ESTs are expected to derive  
from the host plant, no effort was made to eliminate ESTs  
deriving from the pathogen."

## ORIGIN

## Alignment Scores:

Pred. No.: 4.24 Length: 509  
Score: 77.50 Matches: 31  
Percent Similarity: 44.04% Conservative: 17  
Best Local Similarity: 28.44% Mismatches: 34  
Query Match: 15.08% Indels: 27  
DB: 4 Gaps: 5

US-10-089-514-4 (1-103) x BM326074 (1-509)

QY 4 GlnAsnGluLeuGlnArg-----LeuArgAlaGluLeuAspAlaLeuAsp----- 18  
Db 361 CAAAGACAGCTCTCCGCGCAAGACCTCCGACACCGCGGTGCGATCATGATGTC 302  
QY 19 -----GlyThrLeuLeuAspThrValAlaArgArgAlaGlyLeuAspGlyValArgIleAla 36  
Db 301 TTTCTCGGCGCTCGACGACGCTCTTCACGCGCTCGACGCGCCCTTCAC----- 248  
QY 37 ArgTyrLysSerArgHisGlyValProMetMetGlnProGlyArgValSerLeuValLys 56  
Db 247 -----GSCAGAACGCGGACCATCGAC 227  
QY 57 AspArgAlaAlaArgTyrAlaAlaAspHis-----GlyLeuAspGluSerPheLeuVal 74  
Db 226 GAGAACCGCGACTCGACGACCGCGACCGCGCTCGAGTGAATGATCTTTCTCAG 167  
QY 75 AsnLeuTyrAspValIleIleThrGluMetCysArgValGluAspLeuValMetSerArg 94  
Db 166 CGGCTCGACGACGCTCTTCAG-----CGACTCGACGACGCGCTTCACGCCCGA 116  
QY 95 GluSerLeuThrAlaGluAspArgArg 103  
Db 115 ACGGCGACCATCGACGAGAACCGCGA 89

## RESULT 7

BH454175 692 bp DNA linear GSS 12-DEC-2001

LOCUS BOGIX86TF BOGI Brassica oleracea genomic clone BOGIX86, genomic

survey sequence.

ACCESSION BH454175



VERSION	BH454175.1	GI:17639886
KEYWORDS	GSS.	
SOURCE	Brassicica oleracea	
ORGANISM	Brassicica oleracea	
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Eukaryota; Viridiplantae; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosid2; Brassicales; Brassicaceae; Brassica.	
REFERENCE	1 (bases 1 to 692)	
AUTHORS	Town,C.D., Van Aken,S., Uteerback,T., Koo,H. and Fraser,C.M.	
TITLE	Whole genome shotgun sequencing of Brassica oleracea	
JOURNAL	Unpublished (2001)	
COMMENT	Other GSSs: BOGIX86TR Contact: Chris Town	
	TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA. Tel: 301-838-3523 Fax: 301-838-0208 Email: cdtown@tigr.org DNA is from a doubled haploid provided by Tom Osborn. Seq primer: TP Class: sheared ends.	
FEATURES	Location/Qualifiers	
source	1..692 /organism="Brassicica oleracea" /mol_type="genomic DNA" /strain="TO1000DH3" /db_xref="taxon:3712" /clone="BOGIX86" /clone_1b="BOGI" /note="vector: pHD1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHD1 using BstXI linkers"	
ORIGIN		
Alignment Scores:		
Pred. No.:	7.27	Length: 692
Score:	77.00	Matches: 33
Percent Similarity:	42.59%	Conservative: 13
Best Local Similarity:	30.56%	Mismatches: 46
Query Match:	14.98%	Indels: 16
DB:	8	Gaps: 4
US-10-089-514-4 (1-103) x BH454175 (1-692)		
OY	6	GlutenglnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrVal 25
DB	363	CAACTGAACGAGATTGAGGGCTGATCTGCAGAGGCTCGACGGGGTATGACCCGACCGCT 304
OY	26	ArgArgArgGlyLeuAspLeuGlyValArgIleAlaArgTyrLysSerArgHisGlyValPro 45
DB	303	GCTTCACCCGACGAGAGTTT-----AAACCTGGCTGATTGCTTCGACGAACCGCTCCGC 250
OY	46	MetMetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyrAlaAlaAsp 65
DB	249	GAGCTTGAACAGAAAAACACGATCTTCAGAGTGAAGAGGGCTTCA---CTGCTGCCCCG 193
OY	66	HisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIleThrGluMetCys 85
DB	192	CAC-----GACCGAGAGATGAAACGCTCGAGAGACTCCAGAACTTGGAGGCTAGC 142
OY	86	ArgValGluAspLeuValMetSerArgLysLeuThrAla----- 99
DB	141	AGAGAAAGGGGAGAGTGAAGCCGAAATGCGCCCAAGAGCTAATCGTTGTTGCCAG 82
OY	100	-----GluAspArgArg 103
DB	81	ATTGCTTCTCGAGAGGATCGTCGG 58
RESULT 8		
LOCUS	BU915943	913 bp mRNA linear EST 17-OCT-2002
DEFINITION	AGENCECOIRT.10492829 NICHD_XCC_CO1 xenopus laevis cDNA clone	
IMAGE	6642739 5', mRNA sequence.	
ACCESSION	BU915943	

VERSION BU915943.1 GI:24097857  
KEYWORDS EST.  
SOURCE Xenopus laevis (African clawed frog)  
ORGANISM Xenopus laevis  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
Xenopodinae; Xenopus; Xenopus.  
REFERENCE 1 (bases 1 to 913)  
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgarbs@remail.nih.gov  
Issue Procurement: Martha Rebert, Steven L. Klein, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: NCI-CCAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LIML at:  
http://image.liml.gov  
Plate: LIML4228 row: j column: 19  
High quality sequence stop: 666.

FEATURES  
source location/Qualifiers  
1..913  
    /organism="Xenopus laevis"  
    /mol\_type="mRNA"  
    /db\_xref="taxon:8355"  
    /clone="IMAGE:6642739"  
    /tissue\_type="oocytes"  
    /lab\_host="DH10B (phage-resistant)"  
    /clone\_id="NICHD\_XGC\_001"  
    /note="Vector: PCMV-SPTRE; Site 1: NotI; Site 2: SalI;  
    Cloned unidirectionally. Primer: Oligo dT. Average insert  
    size 2.2 Kb. Constructed by Life Technologies."

ORIGIN

Alignment Scores:

Pred. No.:	12	Length:	913
Score:	76.50	Matches:	25
Percent Similarity:	50.00%	Conservative:	25
Best Local Similarity:	25.00%	Mismatches:	39
Query Match:	14.88%	Indels:	11
DB:	5	Gaps:	3

US-10-089-514-4 (1-103) x BU915943 (1-913)

Cy 1 Methrctgttgcgttcacgagcgaatgcaagccactctcacaggctgagctgaccacaaagc 20  
          :::|||||         :::|||||         :::         :::         :::  
Db 508 TTGCTCTAAAATATAAAAGCTGAAGGCCACGCTCTCACAGGCTGAGCTGACCACAAGAGC 567  
          :::|||||         :::|||||         :::         :::         :::  
Cy 21 Leuleuaspthryalargatrgatgile-----AspleugilyValArgile 35  
          :::|||||         :::|||||         :::         :::         |||  
Db 568 GTGCTGGATGGATGAGCAAGAGAGATTCACAGCTACTCAGAGATTACACAGAGAGAT 627  
          :::|||||         :::|||||         :::         :::         |||  
Cy 36 AlaagtTyrlYssSerArghisglYvalPrometmetcInProclYArgValSerleuVal 55  
          :::|||||         :::|||||         :::         :::         :::  
Db 628 ATCAGAATGGCTTCCAGTCAGAGCATGACGTGGAGAGC-----AAATTAAGCCGAA 681  
          :::|||||         :::|||||         :::         :::         :::  
Cy 56 LyasapArAlalaialatgyrYlaialaasPhisiglyLeuasbpluseRphieleuValaen 75  
          :::|||||         :::|||||         :::         :::         |||  
Db 682 CGCAGAGGGCGAGAGAGAGAGGCTGCGCAACACAGAGATGCTCTTAATGCTCTGGAAAC 741  
          :::|||||         :::|||||         :::         :::         |||  
Cy 76 leutyraspyalllellethrjumeCyvArxValglunAspleuValmetSerArsglu 95  
          |||         |||         |||         |||         |||         |||  
Db 742 TTGAGC-----CCAGAGAAATTGCCGGTTATCGGAATCTCTTCAGACGACGAG 789  
          |||         |||         |||         |||         |||         |||

RESULT 9  
CN894351 627 bp mRNA linear EST 04-JUN-2004  
LOCUS CN894351/c  
DEFINITION 010625AANA002796HT (AAVA) Royal Gala 126 DAFB fruit cortex Malus x  
domestica cDNA clone AAVIA002796, mRNA sequence.  
ACCESSION CN894351



VERSION	CN894351.1	GI:48280593
KEYWORDS	EST	
SOURCE	Malus x domestica (cultivated apple)	
ORGANISM	Malus x domestica	
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosid1; Rosales; Rosaceae; Maloideae; Malus.	
REFERENCE	1 (bases 1 to 627)	
AUTHORS	Benning, L., Bowen, J., Crowhurst, R., Gleave, A., Janssen, B., McAtney, S., Newcomb, R., Ross, G., Snowden, K., Walton, E. and Yauk, Y.	
TITLE	HortResearch Apple EST Project	
JOURNAL	Unpublished (2004)	
COMMENT	Contact: Gleave, A. Sequencing Facility The Horticulture and Food Research Institute of New Zealand Ltd 120 Mt Albert Rd, Mt Albert, Auckland, New Zealand Tel: 00 64 09 815 4200 Fax: 00 64 09 815 4201 Email: est@hortresearch.co.nz.	
FEATURES	location/Qualifiers	
source	1..627	
	/organism="Malus x domestica"	
	/mol_type="mRNA"	
	/db_xref="taxon:3750"	
	/clone="AY002796"	
	/tissue_type="fruit cortex"	
	/dev_stage="126 days after full bloom"	
	/clone_11b="(AY01) Royal Gala 126 DAFB fruit cortex"	
	/note="Vector: pBluescript SK(-); Library sequenced by Genesis Research & Development"	
ORIGIN		
Alignment Scores:		
Pred. No.:	10.3	Length: 627
Score:	75.50	Matches: 26
Percent Similarity:	58.33%	Conservative: 9
Best Local Similarity:	43.33%	Mismatches: 20
Query Match:	14.69%	Indels: 5
DB:	7	Gaps: 4
US-10-089-514-4 (1-103) x CN894351 (1-627)		
QY	11 ArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrValArgArgIleAsp 30	
DB	266 AGAGAGGAGGCTCGAGATTCTT---GGGACAGAGCTGTGGCTCCGAGAGGTACTTAC 210	
QY	31 ---LeuGlyValArgIleAlaArgTyrLysSerArgHisGlyValProMetMetGlnPro 49	
DB	209 GAGCTGGGTGCACGC-----CGGACACGAAATCTCTCGGGGTACCGCTCGGCACAGCTT 156	
QY	50 GlyArgValSerLeuValIleAspArgAlaAlaArg---TyrAlaAlaAspHisGlyIleu 68	
DB	155 GACCAAGACAGGCTCGACACGCCGAGAGTGGCCGACCGAGTACCGCGGAATTGTGACTC 96	
RESULT 10		
BM595564/c	529 bp	mRNA linear EST 25-FEB-2002
LOCUS	17000687475785 A.Gam.ad.cDNA.blood1	Anopheles gambiae cDNA clone
DEFINITION	1960049718069 5', mRNA sequence.	
ACCESSION	BM595564	
VERSION	BM595564.1	GI:18891426
KEYWORDS	EST.	
SOURCE	Anopheles gambiae (African malaria mosquito)	
ORGANISM	Anopheles gambiae	
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.	
REFERENCE	1 (bases 1 to 529)	
AUTHORS	Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charlab, R., Collins, F.H., Venter, J.C. and Hoffman, S.L.	
TITLE	Celera Anopheles gambiae EST project	
JOURNAL	Unpublished (2002)	
COMMENT	Contact: Holt R.A.	

	Celera Genomics 45 W. Gude Dr., Rockville, MD 20850, USA Tel.: 2404531151 Fax: 2404534580 Email: HolterA@celera.com Plate: NU010049VS row: J column: 07 Seq primer: M13 Reverse.  location/Qualifiers
FEATURES	
source	1..529 <del>/organism=</del> "Anopheles gambiae" <del>/mol_type=</del> "mRNA" <del>/strain=</del> "KSP-ST (Reduced susc. to Permethrin - std. chromosome)" <del>/db_xref=</del> "taxon:7165" <del>/clone=</del> "19600449718069" <del>/dev_stage=</del> "Adult" <del>/lab_host=</del> "MDH10b" <del>/clone_lib=</del> "A.Gam.ad.cdna.blood1" <del>/note=</del> "Vector: pSport1; Site_1: SalI; Site_2: NotI; whole adult mosquitoes (mixed sex) frozen on liquid nitrogen 24 hours after human blood feeding. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3' clones available through the malaria research and reference reagent resource center ( <a href="http://www.malaria.mr4.org">www.malaria.mr4.org</a> )"
ORIGIN	
Alignment Scores:	
Pred. No.:	9.76                 Length:                 529
Score:	75.00                 Matches:                 22
Percent Similarity:	47.76%               Mismatch:                 10
Best Local Similarity:	32.84%               Indels:                     10
Query Match:	14.59%
DG:	4                         Gaps:                     3
US-10-089-514-4 (1-103) x BMS95564 (1-529)	
OY	2 ThrGtGlnGaNgGluLeuGlNArgLeuArgAlaGluLeuAspAlaLeuAspClYThrlEu 21
Dd	377 ACCCGAGTAATTCACCTTCAAGAGCCTG-----GTGGATGGATTGAAGGCCTTGCT 327
OY	22 LeuASpThrValArGArGARgAlaSpLeuglYValArgllealArGYrLySeSerARg 41
Dd	326 GTGACCgggcgcgcgcgcgccttcgttgccatgsgc-----CACCTGCCAAgCCTC 279
OY	42 HisGLyVALProMetMet-----GInPrcOglYArGVAlSerleuValLYaASParg 58
Dd	278 CACCGAATGTCCGCGTAGACTTTGTGCTCCTCCCGACTGCCAGACTCACCCGTTTCAc 219
OY	59 AlaAlaRgrTYrAlaAlaASP 65
Dd	218 GTTGCGCCCTGCGTTCCAGAG 198
RESULT 11	
BHE54655/C	
LOCUS	BHE54655 644 bp DNA linear GSS 19-FEB-2002
DEFINITION	BOHTELOTF BO 2.3 KB Brassica oleracea genomic clone BOHTE10, genomic survey sequence.
ACCESSION	BHE54655
VERSION	BHE54655.1 GI:18712878
KEYWORDS	GSS.
SOURCE	Brassica oleracea
ORGANISM	Brassica oleracea Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica. 1 (bases 1 to 644) Town,C.D., Van Aken,S., Utecherback,T., Koo,H. and Fraser,C.M. Whole genome shotgun sequencing of Brassica oleracea unpublished (2001) Contact: Chris Town
TITLE	JOURNAL
COMMENT	
	TIR
	9712 Medical Center Drive, Rockville, MD 20850, USA. Tel: 301-838-3523



Fax: 301-838-0208  
Email: [cdtown@tigr.org](mailto:cdtown@tigr.org)  
DNA is from a doubled haploid provided by Tom Osborn

## FEATURES

```

Location/Qualifiers
1..664
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOHTE10"
/clone_lib="BO_2_3 KB"
/note="Vector: pBOS1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into pBOS1 using BstXI linkers"

```

**ORIGIN**

**Alignment Scores:**

Pred. No.:	12.5	644	
Score:	75.00	Matches:	32
Percent Similarity:	43.52%	Conservative:	15
Best Local Similarity:	29.63%	Mismatches:	45
Query Match:	14.59%	Indels:	16
DB:	8	Gaps:	4

US-10-089-514-4 (1-103) X BH654655 (1-644)

Qy	6	GILUEGLINARGLEUARGIAGLUEUAPALALAEASGLYTHLEUUAAPTHVAL	25
	:::		
Db	369	CAACTGAACAGATTGAGGGCTGATCTGCAGAGAGCTCGACGGGTGATATACC	310
			CGCAGCGCT
Qy	26	ARGARGARGILEASPLEUGLYVALARGILEALAAAGTYLYLSERARGHIGLYVAL	45
	:::		
Db	309	GCTCGACGGCGATAGTT-----AAAGCTGGCTGGATTCCTTCGACGAAGTCCGC	256
Qy	46	METMECGLNPROGLYARGVALSERLEUVALIYSAAPARGALAAIARGTYRLAALAAP	65
	:::		
Db	255	GAGCTTGACAGAAAAACACCGATCTCGAGAGTGAAGAGGGCTTCA---CTCGCTGGCAGG	199
Qy	66	HIGGLYLEUASPLUSERPHLEUVALANLEUITYFAAPVALILELTHGLUMECYS	85
Db	198	CAC-----GAGCGAAGAGTAAACCTCGAAGACTCCAGATCTCGAAGTAGC	148
Qy	86	ARGVALGLUASPLEUVALMETSEARGIUSELEUTHALA-----	99
Db	147	AGAGAAAGGGGAGAGTACAGCGCGAALATGGCCGCAAGGCTAATCGTGTGGTCCAGG	88
Qy	100	-----GLUASPARARG	103
Db	87	ATTTCGTTTCGAGGAGCGTCCG	64

RESULT 12	
LOCUS	BH248862
DEFINITION	BH248862 756 bp DNA linear GSS 26-NOV-2001 BOGAM09TF BOGA Brassica oleracea genomic clone BOGAM09, genomic survey sequence.

Fax: 301-838-0208  
Email: [cdtown@tigr.org](mailto:cdtown@tigr.org)  
DNA is from a doubled haploid provided by Tom Osborn

## FEATURES

```

location/Qualifiers
1..756
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGAM09"
/clone_1bp="BOGA"
/notes="Vector: pHOS1, Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers

```

## ORIGIN

**Alignment Scores:**

Pred. No.:	15.2	Length:	756
Score:	75.00	Matches:	20
Percent Similarity:	37.78%	Conservative:	14
Best Local Similarity:	22.22%	Mismatches:	34
Query Match:	14.59%	Indels:	22
DB:	8	Gaps:	1

US-10-089-514-4 (1-103) X BH248862 (1-756)

QY		13	GltLeuAspAlaLeuAspGlyThrValArgArgGileAspLeu---	31
Dd		294	CAAGCCGAGCATTCACAAGAAGATATTACGGTATCAGAAACCCTCGACTTAAG	353
QY		31	-----	31
Dd		354	AAGGATGTGGGCAGACGAACCTCGCGCGCTTTATGTCACACCGAACCCACAGA	413
QY		32	---GlyValArgIleAlaArgTyrllySerArgHisglYalProMetMetGlnProGly	50
Dd		414	GGAGCCACAAAGACCAACCCCATTTCTGCTGCTGCATCACGGAATCGAAGCTATGCCCCGCT	473
QY		51	ArgValSerLeuVallysAsparGalaAlaArgTyrlAlaIsaphisglYleuAspGlu	70
Dd		474	GAACTTAGCCTTAACAAGCTTACCGGGGGTCTMAAGATGCCGACAGACCTAGAGCTCAACAGA	533
QY		71	SerPheLeuValasLeuTyrlAspValIle	80
Dd		534	GATATGCTATTATAACCATTTGGACGCTATC	563

RESULT 13	
CD602770	
LOCUS	312 bp mRNA linear EST 16-JUN-2003
DEFINITION	RK35A2F12.T3 Zebrafish Kidney Marrow cDNA library Danio rerio CDNA clone RK35A2F12 5', mRNA sequence.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 312)  
Song,H.D., Wu,X.Y., Sun,X.J., Zhou,Y., Liu,T.X., Zhang,G.W.,  
Sheng,Y., Chen,Y., Ruan,Z., Jiang,C.L., Fan,H.Y., Look,A.T.,  
Zou,L.I. and Chen,Z.  
Gene Expression Profiling in the Zebrafish Kidney Marrow Tissue  
Unpublished (2003)  
Contact: Chen Z.



Email: zchen@scn.sh.cn  
Seq primer: T3.  
Location/Qualifiers  
source 1..312

/organism="Danio rerio"  
/mol\_type="mrna"  
/db\_xref="taxon:7955"  
/clone="RK354A2F12"  
/dev\_stage="mature"  
/note="Zebrafish Kidney Marrow cDNA library"  
/clone\_1ib="Zebrafish Kidney Marrow cDNA library"  
/note="Organ: Kidney; Vector: pBS-CMV; Site 1: XhoI;  
Site 2: EcoRI; Total RNA was extracted from the kidney  
tissues of mature zebrafish. The poly (A) + RNA fraction  
was separated from total RNA by oligo (dT) cellulose  
chromatography. Library was initially constructed in the  
lambdaZAP Express vector (Stratagene) and in vivo excised  
into pBS-CMV vector."

## ORIGIN

Alignment Scores:  
Pred. No.: 6.93 Length: 312  
Score: 74.00 Matches: 24  
Percent Similarity: 48.15% Conservative: 15  
Best Local Similarity: 29.63% Mismatches: 32  
Query Match: 14.40% Indels: 10  
DB: 6 Gaps: 2  
US-10-089-514-4 (1-103) x CD602770 (1-312)

QY 19 G|YThleuLeuAspThrValArgArgArgIleAspLeuGlyValArgIleAlaArgTyr 38  
Db 1 GGCACGAGCATGTAT-----GGAGTCATGGAGGCGTATTAC 36  
QY 39 LysSerArgHisGlyValProMetMetGlnProGlyArgValSerLeuValLysAspArg 58  
Db 37 CAGAGCTTAAATCTGTCACGCTCTACGCTCCACCACTTCTCACCCTCATCAACCAT 96  
QY 59 AlaAlaArgTyrAlaAlaAsp-----HisGlyLeuAspGlnSerPheLeuValAsnLeu 76  
Db 97 GTAGCCAGATACCGCCGCTCAGTAAAGAGAGCGATGCGATCTTCCTCTCATCATC 156  
QY 77 TyrAspValIleIleThrGlnMetCysArgValGluAspLeuValMetSerArgGlnSer 96  
Db 157 ACCGATGAGTGAATTCAGATATGCGCCAGACAAGAGATGCCATTGTATGTGTCATCT 216  
QY 97 Leu 97  
Db 217 CTT 219

RESULT 14  
LOCUS CD602824 325 bp mRNA linear EST 16-JUN-2003  
DEFINITION RK354A4G12 T3 Zebrafish Kidney Marrow cDNA library Danio rerio cDNA  
clone RK354A4G12 5', mRNA sequence.

ACCESSION CD602824  
VERSION CD602824.1 GI:31784176  
KEYWORDS EST.  
SOURCE Danio rerio (zebrafish)  
ORGANISM Danio rerio

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 325)  
Song,H.D., Wu,X.Y., Sun,X.J., Zhou,Y., Liu,T.X., Zhang,G.W.,  
Sheng,Y., Chen,Y., Ruan,Z., Jiang,C.L., Fan,H.Y., Look,A.T.,  
Zou,J.I. and Chen,Z.  
Gene Expression Profiling in the Zebrafish Kidney Marrow Tissue  
Unpublished (2003)  
Contact: Chen Z.

TITLE  
JOURNAL  
COMMENT State Key Lab for Medical Genomics  
Shanghai Institute of Hematology, Ruijin Hospital Affiliated to  
Shanghai Second Medical University  
197 Rui Jin Road II, Shanghai 200025, P. R. China

Tel: 86-21-64740490  
Fax: 86-21-64743206  
Email: zchen@scn.sh.cn  
Seq primer: T3.  
Location/Qualifiers  
source 1..325

/organism="Danio rerio"  
/mol\_type="mrna"  
/db\_xref="taxon:7955"  
/clone="RK354A4G12"  
/dev\_stage="mature"  
/clone\_1ib="Zebrafish Kidney Marrow cDNA library"  
/note="Organ: Kidney; Vector: pBS-CMV; Site 1: XhoI;  
Site 2: EcoRI; Total RNA was extracted from the kidney  
tissues of mature zebrafish. The poly (A) + RNA fraction  
was separated from total RNA by oligo (dT) cellulose  
chromatography. Library was initially constructed in the  
lambdaZAP Express vector (Stratagene) and in vivo excised  
into pBS-CMV vector."

## ORIGIN

Alignment Scores:  
Pred. No.: 7.29 Length: 325  
Score: 74.00 Matches: 24  
Percent Similarity: 48.15% Conservative: 15  
Best Local Similarity: 29.63% Mismatches: 32  
Query Match: 14.40% Indels: 10  
DB: 6 Gaps: 2  
US-10-089-514-4 (1-103) x CD602824 (1-325)

QY 19 G|YThleuLeuAspThrValArgArgArgIleAspLeuGlyValArgIleAlaArgTyr 38  
Db 1 GGCACGAGCATGTAT-----GGAGTCATGGAGGCGTATTAC 36  
QY 39 LysSerArgHisGlyValProMetMetGlnProGlyArgValSerLeuValLysAspArg 58  
Db 37 CAGAGCTTAAATCTGTCACGCTCTACGCTCCACCACTTCTCACCCTCATCAACCAT 96  
QY 59 AlaAlaArgTyrAlaAlaAsp-----HisGlyLeuAspGlnSerPheLeuValAsnLeu 76  
Db 97 GTAGCCAGATACCGCCGCTCAGTAAAGAGAGCGATGCGATCTTCCTCTCATCATC 156  
QY 77 TyrAspValIleIleThrGlnMetCysArgValGluAspLeuValMetSerArgGlnSer 96  
Db 157 ACCGATGAGTGAATTCAGATATGCGCCAGACAAGAGATGCCATTGTATGTGTCATCT 216  
QY 97 Leu 97  
Db 217 CTT 219

RESULT 15  
LOCUS CP857652 535 bp mRNA linear EST 31-OCT-2003  
DEFINITION psmY0021F03r Agriculture Canada Phytophthora sojae EST project\_gmy  
Phytophthora sojae cDNA clone smY002F03 5', mRNA sequence.

ACCESSION CP857652  
VERSION CP857652.1 GI:38097510  
KEYWORDS EST.  
SOURCE Phytophthora sojae  
ORGANISM Phytophthora sojae

REFERENCE  
AUTHORS Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;  
Phytophthora.  
1 (bases 1 to 535)  
Tyler,B. Not Published  
Unpublished (2003)  
Contact: Tyler B

TITLE  
JOURNAL  
COMMENT Tyler lab  
VBI

1880 Pratt Dr., Blacksburg, VA 24061, USA  
Tel: 540-231-7318  
Email: bmyler@vt.edu



PCR Primers  
 FORWARD: M13 reverse 17mer at 5' end  
 BACKWARD: M13 reverse 17mer at 5' end  
 Plate: 002 row: F column: 03  
 Seq primer: M13 reverse 17mer at 5' end  
 High quality sequence stop: 535.  
 Location/Qualifiers

FEATURES  
 source  
 1..535  
 /organism="Phytophthora sojae"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:67593"  
 /clone="SMY002F03"  
 /issue\_type="mycelium"  
 /cell\_line="P6497"  
 /dev\_stage="mycelium"  
 /clone\_lib="Agriculture Canada Phytophthora sojae EST  
 project\_smy"  
 /note="Vector: pBK-CMV; Site\_1: EcoRI; Site\_2: XhoI"

## ORIGIN

## Alignment Scores:

Pred. No.:	13.6	Length:	535
Score:	74.00	Matches:	25
Percent Similarity:	51.79%	Conservative:	4
Best Local Similarity:	44.64%	Mismatches:	19
Query Match:	14.40%	Indels:	8
DB:	7	Gaps:	3

US-10-089-514-4 (1-103) x CF857652 (1-535)

```

QY      18  AspGlyThrLeuLeuAspThrValArgArgIleAspLeuGlyValArgIleAlaArg 37
          |||||
Db      264  GACGAGAGAGACGCGCTCGAGTACGGCGAGCTGCCGAC--GGAGAGGACTGGCGCGC 320
          |||||

QY      38  TyrIysSerArgHisGlyValPrometMetGlnProGlyArgValSerLeuValIysAsp 57
          ::|||
Db      321  CGACGGGGCCGT-----GTCGCCCGCTCAAGAACCAAGGCAAGTGGCGCTGTCGTGAC 374
          ::|||

QY      58  ArgAlaAlaArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeu 73
          |||||
Db      375  -----CTTCTGACACACGGGCTGCTCGAGGCCACCTG 407
          |||||

```

Search completed: October 5, 2005, 18:37:47  
 Job time : 1132.79 secs



This Page Blank (uspto)







Best Local Similarity 100.0%; Pred. No. 5.6e-66;  
Matches 969; Conservative 0; Mismatches 0; Indels 0; Gaps 0

[illegible]

## RESULT 2

BD093916	LOCUS	BD093916	969 bp	DNA	linear	PAT 27-AUG-2002
DEFINITION	Transformants that produce secondary metabolites modified by a functional group(s) and novel biosynthesis genes.					
ACCESSION	BD093916	BD093916.1	GI:22639504			
VERSION	MO 0123542-A/3.					
KEYWORDS	Streptomyces venezuelae					
SOURCE	Streptomyces venezuelae					
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.					
REFERENCE	Yanai,K., Okakura,K., Yasuda,S., Watanabe,M., Miyamoto,K., Mido,N. and Murakami,T.					
AUTHORS	Transformants that produce secondary metabolites modified by a functional group(s) and novel biosynthesis genes					
TITLE	Patent: WO 0123542-A 3 05-APR-2001;					
JOURNAL	MEIJI SEIYA KAISHA LTD. KOJI YANAI, KORU OKAKURA, SHOHEI YASUDA, MANABU WATANABE, KOICHI MIYAMOTO, TAKESHI MURAKAMI					
COMMENT	OS Streptomyces venezuelae					
	PN WO 0123542-A/3					
	PD 05-APR-2001					
	PF 29-SEP-2000 WO 2000JP006783					
	PR 29-SEP-1999 JP 99P 276314					
	PI KOJI YANAI, KORU OKAKURA, SHOHEI YASUDA, MANABU WATANABE, KOICHI MIYAMOTO, TAKESHI MURAKAMI					
	PI NAOKI MIDO, TAKESHI MURAKAMI					
	PC C12N15/09,C12N5/10,C12P21/02,C07K11/00//C12P21/02,C12R1:645					
FEATURES	Location/Qualifiers					
source	1..969					
	/organism="Streptomyces venezuelae"					
	/mol_type="genomic DNA"					
	/db_xref="taxon:54571"					
ORIGIN						
Query Match	100.0%; Score 969; DB 6; Length 969;					
Best Local Similarity	100.0%; Pred. No. 5,6e-66;					
Matches	969; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1 ATGAGCGGCTTCCCGGACAGGTGTGTGTGCGGGCGACAGCGGGGCGGTGGCGGATGTTCC	60				
DB	1 ATGAGCGGCTTCCCGGACAGGTGTGTGTGCGGGCGACAGCGGGGCGGTGGCGGATGTTCC	60				
QY	61 GCCGGGCGTGGGGGAGGGGGGAGGCGGACGCTGTCGTGCACTGTGACCGCGCGG	120				
DB	61 GCCGGGCGTGGGGGAGGGGGGAGGCGGACGCTGTCGTGCACTGTGACCGCGCGG	120				
QY	121 GGAAGCGGCGGACGCGCTGCTGGTGGGAGGACGTCACCGCGCGGGAGCCGAACTCGGAGCC	180				
DB	121 GGAAGCGGCGGACGCGCTGCTGGTGGGAGGACGTCACCGCGCGGGAGCCGAACTCGGAGCC	180				
QY	181 GCCCTCGGGACGCGGACCTTCGTCCTGCTGCGCTGACACGAGGACGTGGCCCTCAAGGCC	240				
DB	181 GCCCTCGGGACGCGGACCTTCGTCCTGCTGCGCTGACACGAGGACGTGGCCCTCAAGGCC	240				
QY	241 GTGGGCGCCGGTGAACCGGCTCATGCGGCGGGCGCGCGCTGTCGGCGGACACCTGTCGCTC	300				
DB	241 GTGGGCGCCGGTGAACCGGCTCATGCGGCGGGCGCGCGCTGTCGGCGGACACCTGTCGCTC	300				
QY	301 CGAGCGGGCATATGCGCGGAGACTTGGCGAGCCACGCGCCCGGCGTCCAGACAGTGGGCTTC	360				
DB	301 CGAGCGGGCATATGCGCGGAGACTTGGCGAGCCACGCGCCCGGCGTCCAGACAGTGGGCTTC	360				
QY	361 AACCGATGTTGCGCCCGCGCGCGGAGATGACCGGCGGACCGCGTGGCGCGCGTGTCAAC	420				
DB	361 AACCGATGTTGCGCCCGCGCGCGGAGATGACCGGCGGACCGCGTGGCGCGCGTGTCAAC	420				
QY	421 AGGAGCGGGCGCGGCGTCAAGGCGCTTGTGGTGTGTGTGAGAGGGCGGCGGAGGCGCC	480				
DB	421 AGGAGCGGGCGCGGCGTCAAGGCGCTTGTGGTGTGTGTGAGAGGGCGGCGGAGGCGCC	480				



Oy	481	GTACGGCTCA	CGGCGGAGAGACGACGAGCGAGCGGCGACCCAGGCTTGACGAC	540
Db	481	GTACGGCTCA	CGGCGGAGAGAGACGACGAGCGAGCGGCGACCCAGGCTTGACGAC	540
Oy	541	GCCGTGCTCT	CTCTCTTCCGTGGGCTCGGCTCGCCCGGCTTGAGCGTCCGGGCTGTG	600
Db	541	GCCGTGCTCT	CTCTCTTCCGTGGGCTCGGCTCGCCCGGCTTGAGCGTCCGGGCTGTG	600
Oy	601	GCGGGAAGGCA	CCGCGCGCCACACGAGTGTCTGCTCGCTCTTCTGAGCGTCCGGGCTGTG	660
Db	601	GCGGGAAGGCA	CCGCGCGCCACACGAGTGTCTGCTCGCTCTTCTGAGCGTCCGGGCTGTG	660
Oy	661	GCGAGCCCGG	AGGTGTACGAGGACATTCAGCGGCTCAACCCCGGCGGAGCTCGCGGCG	720
Db	661	GCGAGCCCGG	AGGTGTACGAGGACATTCAGCGGCTCAACCCCGGCGGAGCTCGCGGCG	720
Oy	721	CGGGCGTCTG	CGCCGAGCCCTTGCGCTCTTTCGCGGCTGTGTGCGAGCAGCCCGGACGT	780
Db	721	CGGGCGTCTG	CGCCGAGCCCTTGCGCTCTTTCGCGGCTGTGTGCGAGCAGCCCGGACGT	780
Oy	781	GCCGAGCGCC	CGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	840
Db	781	GCCGAGCGCC	CGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	840
Oy	841	AACCTCGAGG	CGGCTTTGCGGGAACTCGCGCGGCTCATGAGGACCGGAGCTCGCGCGGG	900
Db	841	AACCTCGAGG	CGGCTTTGCGGGAACTCGCGCGGCTCATGAGGACCGGAGCTCGCGCGGG	900
Oy	901	CAGGACCACT	GGCAGAGCTGTTCCGCACTTCGACCGGACCGGACCGAGCGAAGCGAGAAG	960
Db	901	CAGGACCACT	GGCAGAGCTGTTCCGCACTTCGACCGGACCGGACCGGACCGAAGCGAGAAG	960
Oy	961	GACCGATGA	969	
Db	961	GACCGATGA	969	
RESULT 3				
AB116234				
LOCUS	AB116234	5251 bp	DNA	linear BCT 03-JUL-2004
DEFINITION	Streptomyces venezuelae pAP, papC, papB, ORFV genes for dehydrogenase, 4-amino-4-deoxychorismate synthase, 4-amino-4-deoxyprephenate protein, complete cds.			
ACCESSION	AB116234			
VERSION	AB116234.1 GI:47846865			
KEYWORDS	Streptomyces venezuelae Streptomyces venezuelae Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.			
SOURCE	ORGANISM			
REFERENCE	1 Yanai, K., Sumida, N., Okakura, K., Moriya, T., Watanabe, M. and Murakami, T. Para-position derivatives of fungal antihelminthic cyclopeptideases engineered with Streptomyces venezuelae antibiotic biosynthetic genes Nat. Biotechnol. 22 (7), 848-855 (2004)			
TITLE	JOURNAL			
REFERENCE	PUBMED 15184904 2 (bases 1 to 5251)			
AUTHORS	Yanai, K., Sumida, N., Okakura, K., Moriya, T., Watanabe, M. and Murakami, T.			
TITLE	JOURNAL			
REFERENCE	Submitted (01-AUG-2003) Koji Yanai, Meiji Seika Kaisha, Ltd., Microbiological Resources and Technology Laboratories, 788 Kayama, Odawara-shi, Kanagawa 2500852, Japan (E-mail:koji_yanai@meiji.co.jp, Tel:81-465-37-5106, Fax:81-465-37-6397)			
FEATURES	Location/Qualifiers			
SOURCE	1..5251 /organism="Streptomyces venezuelae" /mol_type="genomic DNA"			

[illegible]



Query Match 100.0%; Score 969; DB 1; Length 5251;  
Best Local Similarity 100.0%; Pred. No. 2.7e-66;  
Matches 969; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
1 ATGAGCGGGCTTCCCGGACGGTGTGTGCGGCGGACGGGGGGGGGGGGGGGATGTC 60
|
2708 ATGAGCGGGCTTCCCGGACGGTGTGTGCGGCGGACGGGGGGGGGGGGGGGATGTC 2767
|
61 GCCGGGCTGTGCGGAGGCGGGGACCGGACGCTGTGTGTGACCTGTATCCGCCGCG 120
|
2768 GCCGGGCTGTGCGGAGGCGGGGACCGGACGCTGTGTGTGACCTGTATCCGCCGCG 2827
|
121 GGAAGCGCGGACCGCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 180
|
2828 GGAAGCGCGGACCGCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2887
|
181 GCCCTCCGGGACCGGACCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 240
|
2888 GCCCTCCGGGACCGGACCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2947
|
241 GTGGCGCGCGTGAACCGGCTCATAGCGGCGGCGCGGCTGTGTGTGTGTGTGTGT 300
|
2948 GTGGCGCGCGTGAACCGGCTCATAGCGGCGGCGCGGCTGTGTGTGTGTGTGTGT 3007
|
301 CGGACGGGCAATGGCGCGGAGCTTGGCGGCCACGCCCGCGCGTCCAGCACTGTGG 360
|
3008 CGGACGGGCAATGGCGCGGAGCTTGGCGGCCACGCCCGCGCGTCCAGCACTGTGG 3067
|
361 AACCCGATGTTCGCGCGCGCGCGGAGATGACCGGGCGGACCGGTGGCGCGCGGTGC 420
|
3068 AACCCGATGTTCGCGCGCGCGCGGAGATGACCGGGCGGACCGGTGGCGCGCGGTGC 3127
|
421 AGGAGCGGCGCGGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
|
3128 AGGAGCGGCGCGGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3187
|
481 GTACGGCTCATAGCGGCGGAGACGACGACCGGACCGGCGGCGGCGGCGGCGGCGG 540
|
3188 GTACGGCTCATAGCGGCGGAGACGACGACCGGACCGGCGGCGGCGGCGGCGGCG 3247
|
541 GCCGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
|
3248 GCCGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3307
|
601 GCCGCGACCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
|
3308 GCCGCGACCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3367
|
661 GCGACGCCCGGAGGTGTATCGGGGACATTCACGCGGTCCAAACCGCGGCGGCGGCGG 720
|
3368 GCGACGCCCGGAGGTGTATCGGGGACATTCACGCGGTCCAAACCGCGGCGGCGGCGG 3427
|
721 CGGGCGGTGTGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
|
3428 CGGGCGGTGTGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3487
|
781 GCCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
|
3488 GCCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3547
|
841 AACCTCGACGGCGGTCTTTCGGGGAACTTCGCGCGGTCTATGGGACCGGAGCTTCG 900
|
3548 AACCTCGACGGCGGTCTTTCGGGGAACTTCGCGCGGTCTATGGGACCGGAGCTTCG 3607
|
901 CAGGACCACTGTGCGGAGGCTGTTCGCGACCTTCCACCGGACCGACGACGAGGCGGAG 960
|
3608 CAGGACCACTGTGCGGAGGCTGTTCGCGACCTTCCACCGGACCGACGAGGCGGAG 3667
|
961 GACCGATGA 969
|
3668 GACCGATGA 3676
```

RESULT 4  
AF262220  
LOCUS  
DEFINITION  
Streptomyces venezuelae chloramphenicol biosynthetic gene cluster,  
partial sequence.  
AF262220.2 GI:14290414  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Streptomyces venezuelae  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomyces  
Streptomyces venezuelae  
Bacteria; Actinobacteridae; Actinomycetales;  
Streptomyces  
1 (bases 1 to 14159)  
He,J., Magarvey,N., Pirae,M. and Vining,L.C.  
The gene cluster for chloramphenicol biosynthesis in Streptomyces  
venezuelae ISP5230 includes novel shikimate pathway homologues and  
a monomolecular non-ribosomal peptide synthetase gene  
Microbiology 147 (Pt 10), 2617-2829 (2001)  
MEDLINE  
PUBMED  
11571160  
2 (bases 1 to 14159)  
He,J., Magarvey,N.A. and Vining,L.C.  
Direct Submission  
Submitted (01-MAY-2000) Biology, Dalhousie University, 1355 Oxford  
St, Halifax, NS B3H 4J1, Canada  
3 (bases 1 to 14159)  
He,J., Magarvey,N.A. and Vining,L.C.  
Direct Submission  
Submitted (05-JUN-2001) Biology, Dalhousie University, 1355 Oxford  
St, Halifax, NS B3H 4J1, Canada  
Sequence update by submitter  
On Jun 5, 2001 this sequence version replaced gi:10716943.  
FEATURES  
source  
1..14159  
/organism="Streptomyces venezuelae"  
/mol\_type="genomic DNA"  
/strain="ISP5230"  
/db\_xref="taxon:54571"  
/map="adjacent to pabAB"  
233..3511  
/function="essential for chloramphenicol biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="CmlG"  
/protein\_id="AAG21974.1"  
/db\_xref="GI:10716944"  
/translation="MTAGSPGPKTPPRRLAVRDHSFLADGMCDVNPCTAFV  
GTVPEGRVDEPAAMREPVARLGLTRPALDLVEVAEDVVTLDPRPGVEGPAELAD  
LVTVAHNSRPLRIDEAVRLAVRHVDGEYLVKESHVIDLADVDLRLQRAHLRG  
DTPGRIIGRRYHHARIVEQYALPAADPAEYTRALGELVPVGRKGIPTISASVEML  
PREGVTFDELLSAVTGVLPALGGGLVQYVSRMEFATKGYYVEIKELVVRGGLP  
SDTPVEYFAKTHOESIGRPTMSDLDPARAFTSGMRPRVAVVSDTTRPRPDGWRV  
VYVRSARVPEDKFLADRSWPGPPLRLMQYKRGFLSPGAADVDLGLKORIGASGTP  
"3498..6404  
/note="non-ribosomal peptide synthetase; contains  
reductase-like domains near C-terminus"  
/codon\_start=1  
/transl\_table=11  
/product="CmlH"  
/protein\_id="AAG21975.2"  
/db\_xref="GI:14327897"  
/translation="MELPDLITLDFRRHGRGRLATAGRTWYEELDVTSLARID  
AECPRGRVLVAGHTTAAVYVALAAMSHAVTMTMNGPLADRFEEPRVADALLV  
CFERALVGRERAGRLALVAGVGPPTPAAPADGTADDEPARSRVAISIFTSGETGD  
PKLVVHGGLGRLCRSLRLDITPDQVLPALSPASVSELIGLTYAATLVLP  
VADQASWLGVSRYHAAAGCDLALSPVYALDEAARSIRKVFCEALGEGEDYK  
AARVSRVNAVGVTEATVCFSLAELTSTPSIGTVDGPRAVRPDSDGDHATAGTG  
LVIYVDGVALGYAGGSPALENRYGVVDSPPAATDVTISDGGELTYLGRIDEGIKR  
LGRVNLAVGSTLSRHIGREVALVRQDITLILVTAAGEATBEELMRIRDLVWVE  
APDRVLVVALPVTSGKVDKRSALRELLASPAAGHGIGTGDGDAELRRVLDVAVL  
GOEIGPETSIPDAGSSSLAMIQIVKLSDAYGEAEVVEAFAMADYDFAPAFRLRLRG  
EAVAPASAVDTLLRRVETERDALRAELPLRLRDRHREVPVGAADGDEVLITGASGF

## CDS



```

misc_feature      3576..3596
                    /note="Region: adenylation domain A1"
misc_feature      3708..3746
                    /note="Region: adenylation domain A2"
misc_feature      3939..3986
                    /note="Region: adenylation domain A3"
misc_feature      4089..4100
                    /note="Region: adenylation domain A4"
misc_feature      4350..4370
                    /note="Region: adenylation domain A5"
misc_feature      4494..4538
                    /note="Region: adenylation domain A6"
misc_feature      4593..4610
                    /note="Region: adenylation domain A7"
misc_feature      4647..4706
                    /note="Region: adenylation domain A8"
misc_feature      4833..4853
                    /note="Region: adenylation domain A9"
misc_feature      4893..4910
                    /note="Region: adenylation domain A10"
misc_feature      5043..5075
                    /note="Region: thiolation domain"
misc_feature      5343..5375
                    /note="Region: NADH binding site"

```

**ORIGIN**

[illegible]

Db	10699	GCCGTAATCTCTTCCTTCGAGGCTGCGCCTCGCGCCGCTCGAGCTCGAGCTCCGGGCGCTG	10758
OY	601	GCGGCGACCGGACCGCGCGGCCCAACGAGTGTCCTCGCCCTTCGACCCCGTGTGCTCGGC	660
Db	10759	GCGGCGACCGGCACCGCGCGGCCCAACGAGTGTCGCTGCGCCTCGCGCCGCTGTGCTCGGC	10818
OY	661	GGCAGCGCCCGAGGTTGATCGGGGACATCCGACGCTCCAAACCCCGGGCGGCGTCCGCGCG	720
Db	10819	GGCAGCGCCCGAGGTTGATCGGGGACATCCGCGCTCCAAACCCCGGGCGGCGTCCGCGCGC	10878
OY	721	CGGCGCTGTGCGGAGGCGCTCGGCTCTTGGCGGCTGTGTGGGCGAC-----	768
Db	10879	CGGCGCTGTGCGGAGGCGCTCGGCTCTTGGCGGCTGTGTGTGATGCGGACCAACCCGACCGC	10938
OY	769	-----GACCCGGAACGTCGCGGACGCGCCCGGGGCGCGACGCGCCCGGCGATCCCGGG	822
Db	10939	GCCGAGGACCCCGGACCGCGCGGACGACCCCGACCGGACCGACCAACCCCGGCGATCCCGGG	10998
OY	823	GGAATGCGACCGGCGCGCGGGAACCTCGACGCGCGTCTTTCGGGGAACTCCGCGGCTCATGAGGA	882
Db	10999	GGAATGCGACCGGCGCGCGGGAACCTCGACGCGCGTCTTTCGAGGAACCTCGCGGCTCATGAGGA	11058
OY	883	CCGGAGCTCGCGGCGGGGCGGACGACACTGTGCGAGAGCTGTTCGCGACCTTCCACCGCACCC	942
Db	11059	CCGGAGCTCGCGGCGGGGCGGACGACACTGTGCGAGAGCTGTTCGCGACCTTCCACCGCACCC	11118
OY	943	GACGACGAAAGCGGAGAAAGGACCGGATTA	969
Db	11119	GACGACGAAAGCGGAGAAAGGACCGGATTA	11145

RESULT	5
A48324	LOCUS
A48324	DEFINITION
A48324	ACCESSION
A48324.1	VERSION
G1:2302117	KEYWORDS
GI:2302117	SOURCE
Strepomyces pristinaespiralis	ORGANISM
Bacteria; Actinobacteriae; Streptomycetales;	
Streptomycetaceae; Streptomyces.	
(bases 1 to 888)	
Blanc, V., Thibaut, D., Bamas-Jacques, N., Blanche, F., Crouzet, J.,	
Barriere, J., Debussche, L., Famechon, A., Paris, J. and	
Dutric-Rosset, G.	
STREPTOMYCETES AND METHOD FOR PREPARING SAME BY MUTASYNTHESIS	
PATENT: WO 9601901-A 2 25-JAN-1996;	
RHONE POULENC RORER SA (FR)	
Other publication AU 2891295 960209	
Other publication FR 2722210 960112.	
Location/Qualifiers	
1..888	
/organism="Streptomyces pristinaespiralis"	
/mol_type="unassigned DNA"	
/db_xref="taxon:38300"	
ORIGIN	
Query Match	27.9%; Score 270.8; DB 6; Length 888;
Best Local Similarity	58.8%; Pred. No. 1.5e-12;
Matches	486; Conservative 0; Mismatches 337; Indels 3; Gaps 1;
10 TTCCCGGAGAGCTTCGTCTGGCGGCAGCGGGGCGGTGGCGGCAAGTT---CGCCGG 66	
19 TTCGGGGTGTGTGTGTGGTGCGGGGCCGGTGGTGGTGGCCGATGTTACCCACTGG 78	
67 CTGCTGGCGGAGGAGCGGCACCGCAGCCTCGTGCACCTCGTACCTGCAGCGCGCGGAGCG 126	
79 CTGATGCTTCGAGGGGTGCGCGGTACCTGAGCTGACGTGACC CGGGGCGGTGCGCGGAC 138	
127 CCGGAGCGCTTGCTGTGTGGCGAGCTGACCGCGCGCGGGGCGGAATTGCGGCGGCTTC 186	



Db 139 GGGGTCGGAGTGTGTCGGGCTGATATGTCGGGCGCGGGGCGGAGCGGTCGGCGCGT 198  
QY 187 CGGAGCGCGGACCTGCTGCTGCTGCGCCGTACAGAGGACGTGGCCCTCAAGGCGTGGCG 246  
Db 199 GCGGCGCGGACCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 258  
QY 247 CCGGTGACCCGGCTCATGCGCGCGGCGCGCTGCTGCGGACACCTGTCCTTCGGAGC 306  
Db 259 GTGCTGGCGGGGGTGAATGCGCGCGGTGCGGTGCTCCGGACACTTGTGTGGTCAAGAGC 318  
QY 307 GGCATGGCGCGGAGCTCGGGGCGCAAGCCCGCGCGCTCCAGACGTCGGGCGCTCAACCG 366  
Db 319 CGGATGCGCGGCGGAGCTGCTGTAAGCGCGCGCGGAGCTGAGCGGTGGAGCTGAACCG 378  
QY 367 ATGTTGCCCCCGCGCGCGCGCATGACCGGCGGACCCCTGGCGCGCGCTGATCACCAGGAC 426  
Db 379 ATGTTGCCCCCTTCGCTGGGCTTTCAAGGGGCGCGGTGGCGCGGTGGTGGTGGTGGTGG 438  
QY 427 GGGCGGGGCTCACGGGCTGTCGGGCTGTCGAGGGCGCGCGGCGGAGCGCCGTACCG 486  
Db 439 GGGCGCGGTGTGGGGGCTGTGTGAGCTGTGGCGCGGTGGGGGGCGCGGTGTGTGAG 498  
QY 487 CTCACGCGGAGAGACGACGACCGGACGACCGCGCGCACCGAGCCCTGAGCGACGCGT 546  
Db 499 ATGCGCGCGCGGCGGACGAGAGCTGACCGCGCGACGAGCGCGCACGATGCGCG 558  
QY 547 CTCCTCTCTTCGAGGCTGCGCTGCGCCGCTCGGCGCTCGACGTCCGAGGCGCTGGCGCG 606  
Db 559 GTGCTGGCTTTCGGGCTGGGGCTGGGTGAGCTGTGCGGTGAGCGTGGGGGGCGCTGGGGAG 618  
QY 607 ACCGACCGCGCGCCCAACGAGTGTCTGCGCTCTGCGCCGCTGTGTCTGGCGCGAGC 666  
Db 619 AGTCCCCCGCGCGCGCATGTGCGAGTGTGCGCTGCGCGCGGATCGCGCGGAGC 678  
QY 667 CCGAGGTGATCGGGGACATCCAGCGGTCCACCCCGCGCGCGCTCGCGCGCGCGCGCG 726  
Db 679 CCGAGGTGATTTTTCAGATCCAGCGCGCGCGCACCCCGCGCGCGCGCGCGCGAGGCG 738  
QY 727 CTCGCGAGGAGCCTGTGCGCTCTTCGCGCGCTGTGTGCGAGCACCGCGGACCGTGTGCGAG 786  
Db 739 CTGGGCGCGGCGCTGTGTGGGCTGGGGCAGCGCGTGAAGGGGCGACGAGAGAGCTTC 798  
QY 787 GCGCGCGGCGCGCGCGGACCGCGCGCGCGCATCCGCGGGGATGCGAGC 832  
Db 799 GCGCGCTGTTCGCGGAACTGCGCGGTGTGTGGGCGAGCAGCGCG 844

RESULT 6  
ARI98354 888 bp DNA linear PAT 20-APR-2002  
LOCUS ARI98354 Sequence 2 from patent US 6352839.  
DEFINITION ARI98354  
ACCESSION ARI98354.1 GI:20248203  
VERSION  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 888)  
AUTHORS Blanc,V., Thibaut,D., Bamas-Jacques,N., Blanche,F., Crouzet,J.,  
Barriere,J.-C., Debussche,L., Famechon,A., Paris,J.-M. and  
Dutric-Rosset,G.  
TITLE Streptogramins for preparing same by mutasynthesis  
JOURNAL Patent: US 6352839-A 2 05-MAR-2002;  
FEATURES Location/Qualifiers  
source 1..888  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN  
Query Match 27.9%; Score 270.8; DB 6; Length 888;  
Best Local Similarity 58.8%; Pred. No. 1.5e-12;  
Matches 486; Conservative 0; Mismatches 337; Indels 3; Gaps 1;

QY 10 TTCGCCGAGCGTCTGCTGCGCGGACCGGGCGGTGGCGGATGTT--CGCCGG 66  
Db 19 TTCGGCGT 78  
QY 67 CTCCTCCGGAGGCGGCGAGCGCGACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 126  
Db 79 CTGTGTCTTTCGGGGGTGGCGGTGACTGTGACTGTGAATGTGAGCCGGGCGCGGTGGCGGAC 138  
QY 127 CCGAGCGCTGCTGTGTGGGACGTCACCGGCGCGGGGCGCGGATCTCGCGCGCGCTTC 186  
Db 139 GGGGTGCGGGGTGTGTGCGCGGTGATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 198  
QY 187 CGGAGCGGACCTGCTGCTGCTGCGCGGTACAGAGACGTGGCCCTCAAGGCGTGGCG 246  
Db 199 GCGGCGCGGACGTTGT 258  
QY 247 CCGGTGACCCGGCTCATGCGCGCGGCGCGCTGCTGCGGACACCTTGTCTTCGGAGC 306  
Db 259 GTGCTGGCGGGGTGATGCGCGCGGTGCGGTGCTCCGGACACTTGTGTGGTCAAGAGC 318  
QY 307 GGCATGGCGCGGAGCTCGGGGCGCAAGCCCGCGCGCTCCAGACGTCGGGCGCTCAACCG 366  
Db 319 CGGATGCGCGGCGGAGCTGCTGTAAGCGCGCGCGGAGCTGAGCGGTGGAGCTGAACCG 378  
QY 367 ATGTTGCCCCCGCGCGCGCATGACCGGCGGACCCCTGGCGCGCGCTGATCACCAGGAC 426  
Db 379 ATGTTGCCCCCTTCGCTGGGCTTTCAAGGGGCGCGGTGGCGCGGTGGTGGTGGTGGTGG 438  
QY 427 GGGCGGGGCTCACGGGCTGCTGCGGCTGTCGAGGGCGCGCGGCGGAGCGCCGTACCG 486  
Db 439 GGGCGCGGTGTGCGGGGCTGTGTGAGCTGTGTGCGCGGTGGGGGGCGCGGTGTGTGAG 498  
QY 487 CTCACGCGGAGAGACGACGACCGGACGACCGCGCGCACCGAGCCCTGAGCGACGCGT 546  
Db 499 ATGCGCGCGCGGCGGACGAGAGCTGACCGCGCGACGAGCGCGCACGATGCGCG 558  
QY 547 CTCCTCTCTTCGAGGCTGCGCTGCGCCGCTGTGGGTGAGCGTCCGAGGCGCTGGCGGCG 606  
Db 559 GTGCTGGCTTTCGGGCTGGGGCTGGGTGAGCTGTGCGGTGAGCGTGGGGGGCGCTGGGGAG 618  
QY 607 ACCGACCGCGCGCCCAACGAGTGTCTGCTGCTCTGCGCGCTGTGTCTGGCGCGAGC 666  
Db 619 AGTCCCCCGCGCGCGCATGTGCGAGTGTGCGCTGCGCGCGGATCGCGCGGAGC 678  
QY 667 CCGAGGTGATCGGGGACATCCAGCGGTCCACCCCGCGCGCGCTCGCGCGCGCGCGCG 726  
Db 679 CCGAGGTGATTTTTCAGATCCAGCGCGCGCGCACCCCGCGCGCGCGCGCGAGGCG 738  
QY 727 CTCGCGAGGAGCCTGTGCGCTCTTCGCGCGCTGTGTGCGAGCACCGGACCGTGTGCGAG 786  
Db 739 CTGGGCGCGGCGCTGTGTGGGCTGGGGCAGCGCGTGAAGGGGCGACGAGAGAGCTTC 798  
QY 787 GCGCGCGGCGCGCGCGGACCGCGCGCGCGCATCCGCGGGGATGCGAGC 832  
Db 799 GCGCGCTGTTCGCGGAACTGCGCGGTGTGTGGGCGAGCAGCGCG 844

RESULT 7  
A48323/c 2888 bp DNA linear PAT 07-MAR-1997  
LOCUS A48323 Sequence 1 from Patent WO9601901.  
DEFINITION A48323  
ACCESSION A48323  
VERSION A48323.1 GI:230216  
KEYWORDS  
SOURCE Streptomyces pristinaespiralis  
ORGANISM Streptomyces pristinaespiralis  
REFERENCE 1 (bases 1 to 2888)  
AUTHORS Blanc,V., Thibaut,D., Bamas-Jacques,N., Blanche,F., Crouzet,J.,  
Barriere,J.-C., Debussche,L., Famechon,A., Paris,J. and  
Dutric-Rosset,G.  
TITLE STREPTOGRAMINS AND METHOD FOR PREPARING SAME BY MUTASYNTHESIS







QY	727	CTCCGCCAGGCGCCCTGCTCTCTGCGCCGCTGGTCGGCGACGACCCCGACCGTTCGCCAG	786
Db	1098	CTGGGCGCGCGCGCTGGTCGGCGCTGGCGGACGAGCGCGCTGCAGAGGGGGGACGAGGAGAAGACTTTC	10399
QY	787	GCCCCCGGCGCGCGCGACGCCCCCGCGGCATCCCGGAGATCGACGACG	832
Db	1038	GCCGCGCTGTTCGCCGACACTGCGCGTGTCTGTCGGCGGACGACGCGCG	993
RESULT 9	SPU60417	4740 bp	DNA linear BCT 07-MAR-1997
LOCUS	SPU60417/c		
DEFINITION	Streptomyces pristinaespiralis 4-dimethylamino-L-phenylalanine precursor biosynthesis (papa, papC, papB, papM) genes, complete cds.		
ACCESSION	U60417		
VERSION	U60417.1		GI:1575335
KEYWORDS			
SOURCE			
ORGANISM	Streptomyces pristinaespiralis		
REFERENCE	Streptomyces pristinaespiralis		
AUTHORS	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
TITLE	1 (bases 1 to 4740)		
JOURNAL	Blanc, V., Gil, P., Bamas-Jacques, N., Lorenzon, S., Zagorec, M., and Thiabaut, D.		
MEDLINE	Identification and analysis of genes from Streptomyces pristinaespiralis encoding enzymes involved in the biosynthesis of the 4-dimethylamino-L-phenylalanine precursor of pristinamycin I		
REFERENCE	Mol. Microbiol. 23 (2), 191-202 (1997)		
AUTHORS	97197164		
TITLE	2 (bases 1 to 4740)		
JOURNAL	Blanc, V., Gil, P., Bamas-Jacques, N., Lorenzon, S., Schlemminger, J., Biesch, D., Blanche, F., Debussche, L., Crouzet, J. and Thiabaut, D.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (11-JUN-1996) Recherche Pharmaceutique, Rhone-Poulenc Roter, 13 quat Jules Guesde, B. P. 14, Vitry sur Seine cedex 94403, France		
FEATURES			
source	location/Qualifiers		
gene	1..4740		
CDS	/organism="Streptomyces pristinaespiralis"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:38300"		
	68..2227		
	/gene="papa"		
	68..2227		
	/gene="papa"		
	/function="p-aminobenzoate synthase"		
	/codon_start=1		
	/transl_table=1		
	/product="Papa"		
	/protein_id="AAC44866.1"		
	/db_xref="GI:1575336"		
	/translation="MRTVITLLIDNYDSFTYNLFOMLAEVNGAALVVRNDRTFQQA LAPDFNVAVSPGHPADITDGLSRVVTENDPLTGVCGLHQLCLAGAAVVAH APEPHGRTSDIRHGGGLFANIPSPITLVAYHSLIVROLPADLRATADGQMLAV AHRLLPRGVOHFHESISSEHGHEMLANFRDLSRAAGHPHPTERTAPAPAPAP ADAPAPAPGCEVRLAHREYACVPDADAEPALFADAPARWIDSSRYEGLARFTLL GARGPVGEOQTYTVADRAVAVKDGSGGERTRPETLLFDHLEHEILAAALPITGLPEFEF NLGVGYLGIETKADSGEDAHREGELPDGAMFPADRMALDHEGRLMALLSESTRBP ATDAERARWLTDARLTATTPARPPFTLLPDQQLPADIVYRHSLPRYREIVEGRRL IDDEYETEVCLTNMLRVGARIDPLTAYRALRTVAPAYAVIOPGPAVIVSSPERLL RIGADWAESKSPITKGTBRPRAGPADODAASVAAAKDSENMITYDIVRNDGQC DICGVHYPGLFEVETVATVHOVLSVTGRGLADVSRPRAAATPGGSMGAPRVKTM QFIDRLKRPGRGVSGALGYFALSGAADLSIVRTIVATEBAATIGVAGVVALSDPD DEVEMILIKQOTVITLAALROAHGATASDREHLASLR"		
	complement (2486..3376)		
	/gene="papC"		
	complement (2486..3376)		
	/gene="papC"		
	complement (2486..3376)		
	/gene="papC"		
	/function="dehydrogenase"		
	/codon_start=1		

[illegible]



QY	427	GGGCGGGGCGTCACGGGCGCTGGGCGCTGTCGAGGGCGGGCGGCGGACCGGTACGG	486
Db	2938	GGGCGGGGCTGTCGGGGCCCTGGTGGAGCTGTGGCCGGGTGGGGGCGCCGGGTGTGGAG	28797
QY	487	CTCAGCGGAGAGAGACGACCGGACGACCGGCGGCGCAACCAGGCGCTTCAGCGACGCGTG	546
Db	2878	ATGCCGCGGCGGGCGGACGACGAGCTGACCGGCGCGGACGAGGCGCGCACGAGATGCCGCG	2819
QY	547	CTCCCTCTCTTTGAGGCTGGCCCTGCGCCCGCTCGGCGTTCGACGTTCGGGCGCTTGCGGCG	606
Db	2818	GTGCTGGGCGCTTCGGGCTGGGCGCTGGGTGAGCTGTTCGTGGACGTGGGGGCGCTGCGGGAC	27598
QY	607	ACGGCACCGCGCGCCCCACAGGCTGTGTCGCGCCCTCMGGGCCGHTGCTCTGGGGGCGACG	666
Db	2758	AGTGGCCCCGCGCGCGCATCTGGCGATGTGGGCTGTGGCGCCCGGATGCGCGGGGACG	26999
QY	667	CCCGAGGTGTACGGGAGCATCCAGCGGTCACAAACCCCGGGGCGGCTCCGCGCGCGGACG	726
Db	2698	CCGGAGGTGTATTTTGGACATCCAGGCGCGCAACCCCGGCGGCGCGCGCGCGGAGGCG	26399
QY	727	CTCGCCGAGGCCCTGCGCTCTCTTGCGCGCGCTGTGCGCGACGACCCGAGACCTGTCGAC	786
Db	2638	CTGGGCGCGCGCTGTGTGCGGCTGGGCGAGGCGCGTCGAGAGGGGCGACGAGAGCGTTTC	25797
QY	787	GCCCCCGGGCGCGCGACCGCCCGG	812
Db	2578	GCGCGCTGTTTCCCGAATCGCGG	2553

gene  
CDS

RAAAVQAPPPADSPILLDDPPASRRPPAPVSGPPVAPAPPPAPPPAPPPAPPP  
 ALTQAOAQOTQARSQARQAQALALQALQALGRAPRRPAPRONHNTQTPRPAQOATQOTPT  
 PQAQAPQTAQTAQOATQOTQAOQAQOATQOTQOTAPRETQTAQTAQOATQOTQTA  
 PRRPAPSGAGAGSGSGPKPKASLPAPDAAPAPQLPAPPLAAAPAPPPAPSGATRV  
 PGPSAAPPPAPPPAPPPAPPPAPPPAPPPAPPPAPPPAPPPAPPPAPPPAPPPAPPP  
 PISGVSIVVAMPYNTKYTYTGGLPLDMETGAIGAVVVLVSGSDCNLARCLAAEPOM  
 ARKSLPPEAAPPCVSPPEYFGDPBAHGMNTPVGGMLFEGCALIGGSSFSLDSRHPTT  
 PAEGSPD  
 4856. .4861  
 complement (4879. .4884)  
 complement (8315. .8320)  
 9469. .10167  
 /gene="U11"  
 9469. .10167  
 /gene="U11"  
 /note="complexes with gh; involved in membrane fusion"  
 /codon\_start=1  
 /product="virlion glycoprotein L"  
 /protein\_id="AAU88066.1"  
 /db\_xref="GI:52843275"  
 /translation="MANVIRGVLLVLAASLPKPCSPNYLLISPPASSTTEYLQISVAAYV  
 ADVLACGLIDPADQVTRVYRSLSDYRIGCIFLRHPCGLDITVLMDSQARAYV  
 NPFLLTAGCLELUSHLFPDNLLETTRALYKEVLAVASRGAASDPVPPGCVSS  
 EKSRRSTRCIGRTPGFIDEPRLRSPISAPGDEAAPOFSPPAVTSSAPARRRRRPPR  
 KPRRRSTERSRRPR"  
 10235. .10990  
 /gene="U12"  
 10235. .10990

RESULT	10
LOCUS	AY174813                150715 bp      DNA          linear          VRL 04-OCT-2004
DEFINITION	Cercopithecine herpesvirus 2, complete genome.
ACCESSION	AY174813
VERSION	AY174813.1 GI:52843273
KEYWORDS	.
SOURCE	Cercopithecine herpesvirus 2 Cercopithecine herpesvirus 2 Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Alphaherpesvirinae. 1 (bases 1 to 150715) Tyler,S.D., Peters,G.A. and Severini,A. Complete genome sequence of Cercopithecine herpesvirus 2 (SAB) and comparison with other Simplexviruses Virology (2004) In press 2 (bases 1 to 150715) Tyler,S.D., Peters,G.A. and Severini,A. Direct Submission Submitted (11-AUG-2004) Health Canada, National Microbiology Laboratory, 1015 Arlington St., Winnipeg, Manitoba R3E 3R2, Canada
JOURNAL	Location/Qualifiers
REFERENCE	1. .150715 /organism="Cercopithecine herpesvirus 2"
AUTHORS	/mol_type="genomic DNA"
TITLE	/db_xref="taxon:10317"
JOURNLT	/note="SAB"
FEATURES	1. .236 /note="a sequence"
source	237..9207 /note="terminal copy of long repeat region"
	2178..4669 /gene="RL2"
	join(2178..2225,2448..3020,3398..4669)
	/gene="RL2"
	/note="multifunctional regulatory protein"
	/codon_start=1
	/product="immediate early protein ICP0"
	/protein_id="AAU88065.1"
	/db_xref="GI:52843274"
	/translation="MEPSRGIGAPGRSSAQAAGDALDEELMLSRFYPPDSDDSDVDNDGGRBPDADTDALICGPASGSPRAPSPPRVCATCTERICTOICAAPFCLRFCTPCIKTWLPMKNSPLCNNAVYILIVGXKDGSGISTIPINDPRTAE AEEAVRGTAIDFTWTRHLPGEAAPASVTGGRTVALSPARKGQPAFGGAARAP
gene	
repeat_region	
misc_feature	
CDS	

polyA\_signal  
polyA\_signal  
gene  
CDS  
gene

```

"note"= "proposed start codon is GTG"
/codon_start=1
/translation_except= "(pos:10235..10237,aa:Met)"
/product= "uracil-DNA glycosylase"
/protein_id= "AAU88067.1"
/db_xref= "GI:52843276"
/translation= "WTPFSAPRPLGLDDTTPAGRPSPSLMDPERRRFLYGDAMRPILLE
BELAPLTLARLAEIERKRAEAVLPREDVSWTRCTPDEKRVITQDPYPHQ
AHGLAFAVRGPGRAPSLNLTIAVRNCCPEATMSGHGLEKARGVILLNTLVPR
RGEPSHAOIGMDREVGVSIVRLTAASRPGLVMLMGHAQSAIRPPRHRYVLTVLHP
SPLSRPFSGCRHFCLANQYLERSIAPIDMET"
11076..11081
11076..11081
11130..11813
11130..11813
/gene= "UL3"
11130..11813
/gene= "UL3"
/note= "colocalizes with ICP22 and UL4 in small, dense
nuclear bodies"
/codon_start=1
/product= "UL3"
/protein_id= "AAU88068.1"
/db_xref= "GI:52843277"
/translation= "MSNVAGCAVPSTLAIVLSMGAFAADPHDQPTLRKISPAEA
PYGNVPRADDRAAPGCIARPAPPPHYAEPITMWVSVDLEGRQLDTTRKDLRS
NLALICTSSSSGAGAPAPPARIGPREPSHSLSIQMFVLCQKPDATRNDQLRA
DVASEKKPKYYTRSDGRISPAIPVFVHEFSDAPVHRDNVIAVVRERRSTRILA
PIASDEDPMQ"
11809..11814
complement(11812)..11837)
complement(11897)..12511)
/gene= "UL4"
complement(11897..12511)
/gene= "UL4"
/note= "colocalizes with ICP22 and UL3 in small, dense
nuclear bodies"
/codon_start=1
/product= "UL4"
/protein_id= "AAU88069.1"
/db_xref= "GI:52843278"
/translation= "MSASASSATTIVSLHGATSACVLPDAEOVVCAPESGRRAIAS
RGCLRHDLSGAVVVVROTGVILWMDCRTEFCAYRFVGRSEORLEAKMTTLLCAY
FPDSVSSSTRRESARSPTAGIATIYWGDSIIYITTYTGSPPEGEGABPRLLPPPBGD
APDAFSPVAVSPASKEATAADAILVENKEIQLSFTLFGPCDA"
complement(12516..15163)

```



gene

complement (12536, .15163)

/gene="UL5"

/codon\_start=1

/product="component of helicase-primase complex"

/protein\_id="AAU88070.1"

/db\_xref="GI:52843279"

/translation="WTTVGEASHGASASPPPEQPPPSARRDPAFLNFTSMHGIQPI  
IERVELRGLRIEPRVPRPQWFRDVALLETPADLPKEPFAVYLITNGAGSKSTC  
VQTLNEVLDVVATGATRTIAAONMYKLSGALSGSPINTIPIEFGFNHYNQAQGGP  
VALANSPASVEDLQDRDLTYWEVIAIDITRAVAAAGGTGEGSEFPALAEETML  
AGSLITRLAIPAFHGLAPAFTRSNITVIVDEALDRHLITLNAVYCOMTMNLHYTPQVA  
AGRLVLYCVSGSPOTLASBESTEHQKRCISROENVLYILINTRLREVLPRSM  
ALFINKKCVHEFGNLMKALBYGLPTEEMQVDRFVPESTINPAPLPGWTRLP  
SSHKVSAVMARLHAHLKARVEGEVVFLEVLTFVSKRVEERELTHQPALTEVM  
IAANGRIINTVYSSODODAGRYRFEYHKKOOLVAVRDVYVLSQALVATTRKLVY  
GEGFNGRSEAVLRDSDPFTKQGEMLVFAFVPLSRILPGLIHFVNLIRPGIDAK  
RVARERNQALITTELIALGASGABGAAPPAAGEAPDPKTLTGAAADADPEPD  
LDVYRAGASBQEDLPYCHYPEQPDITTAAYHAQPGILKRAFLSRHILBLEGEAE  
GAPFTSYVDNVTFRCGEVWISGMRGMLVALQDTSYTLGFTYARVYARAEVRRH  
AAAGLAELLESPLPYVVLRODHGFMASVNNVNSEFVSVEARLMAATADYGISM  
LAMTITRSGQSLDKVAICFTPGNRLNSAVVASRTTSEFLMNINPLRERHED  
VTSGLSLARPELVLY"

15162, .17198

/gene="UL6"

15162, .17198

/gene="UL6"

/note="role in DNA cleavage/packaging"

/codon\_start=1

/product="capsid protein"

/protein\_id="AAU88071.1"

/db\_xref="GI:52843280"

/translation="MAERHPRGLRPPAMPDSCGPRPADDDGVQAHPPTTLMFREALS  
GLTGGEGGVNVRSSSEAAITROLGAVPFALNATNRHLEADMHRVHAARLEBE  
RLVRYRGAPADVAAPVAREVDTTQRLRASILDPARGLADCAQCGSGSGSY  
IDMLRVCGLGAPDVAAPVAREVDTTQRLRASILDPARGLADCAQCGSGSGSY  
PDSATLADYRCVLFYHPRRGVYRDPDAGRGSCVYLAMPVAGORLLPDSMQL  
SHEIYALRLREHARVCVRNNAASKYVILGKNDGRDPAGAAVARSALGBEDDASA  
GSAARLYALILINMGKMHIGDINTVAYVYIDEGMGLMSATVDPALPFGSAAAG  
RAAGDAPGPRQOOLROAFOTSVVNNINGMLEGYNILPFTIERLENTDILATOLR  
DDLLRLRARGALDRBORADIGRENGGAAADPDADVDIIVSKAMDNDNYAN  
SFOHQVNSYADDLERLSRLMEHLIRFKTSRHTNNGRETAIFYSGAIALFVAY  
LEHARAPLPGALITGPGSVLGEERLMALIFKTRVQVYITDLALFVAADVRAARR  
LEPAHRAAPRGGRGRSPPVDAPGARGADVAVRGDHPRLRGRGGRBP"

17149, .18039

/gene="UL7"

17149, .18039

/gene="UL7"

/note="role in DNA cleavage/packaging"

/codon\_start=1

/product="capsid protein"

/protein\_id="AAU88072.1"

/db\_xref="GI:52843281"

/translation="MAAPSGGEETAAALRQAVAGDRDLVGAEALISHQALLRMACE  
VQAAGRGDVRPATVGSRVVDYTPRLRLFPALDDGPALITASBDYFARCGQAHRGF  
AAVAVLGDHYVSLCVPLVLPRLLSYRPERLEDPLALMYLBNCRPAHTASA  
FKISGMIAAVERSPDRVRCILRSCOMLNTLMVMHVEFDPRYVLPKCMAR  
YLLDINPEPILIALFRAPASRFLRPVPGTINLECAVNPQVLSGMAADBPAGIA  
AMWASDAPRTQALAFYRFV"

18062, .18067

complement (18205, .18210)

complement (18213, .20489)

/gene="UL8"

complement (18213, .20489)

/gene="UL8"

/codon\_start=1

/product="component of helicase-primase complex"

/protein\_id="AAU88073.1"

/db\_xref="GI:52843282"

/translation="WGEDGLAWAEBSVCATVLYAAMPRTTRDRILHALYILCRDARG  
DAVARAEVAVASSELQAFYFGELISAAGAAVAAARAAASPAALLETINPALARLY  
ASALALAEVGVGAPLFPVARGDRPGLVATLEAAGPRVPRGAALGVQAVT  
DEPLAAHYAAVPAARLAWARLAAIOTPCOASASIALVRIDGTAPAEVYTLFPP

polYA\_signal

polYA\_signal

gene

CDS

gene  
complement (20544...23183)  
/gene="UL3"

Query Match	13.3%	Score	128.4	DB	14	Length	150715
Best Local Similarity	47.5%	Pred. No.	0.012				
Matches	381	Conservative				Indels	0
						Gaps	0

[illegible]



RESULT 11  
AY714813/c 150715 bp DNA linear VRL 04-OCT-2004  
LOCUS AY714813  
DEFINITION Cercopithecine herpesvirus 2, complete genome.  
ACCESSION AY714813  
VERSION AY714813.1 GI:52843273  
KEYWORDS  
SOURCE  
ORGANISM  
Cercopithecine herpesvirus 2  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Alphaherpesvirinae.  
1 (bases 1 to 150715)  
2 (bases 1 to 150715)  
REFERENCE  
AUTHORS Tyler,S.D., Peters,G.A. and Severini,A.  
TITLE Complete genome sequence of Cercopithecine herpesvirus 2 (SAB) and  
comparison with other simlexviruses  
JOURNAL Virology (2004) In press  
AUTHORS Tyler,S.D., Peters,G.A. and Severini,A.  
TITLE Direct Submission  
JOURNAL Submitted (11-AUG-2004) Health Canada, National Microbiology  
Laboratory, 1015 Arlington St., Winnipeg, Manitoba R3E 3R2, Canada  
FEATURES  
source  
1. 150715  
/organism="Cercopithecine herpesvirus 2"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10317"  
/note="SAB"  
1. 236  
/note="a sequence"  
237..9207  
/note="terminal copy of long repeat region"  
2178..4669  
/gene="RL2"  
join(2178..2225,2448..3020,3398..4669)  
/gene="RL2"  
/note="multifunctional regulatory protein"  
/codon\_start=1  
/product="immediate early protein ICP0"  
/protein\_id="AAU88065.1"  
/db\_xref="GI:52843274"  
/translation="MEPSRGLGAPGRSSAQAADALDELLMLSRFFPSSDDSD  
VDDDDGGRPRPADDTDAELCEGPASGRPRAPPPRECAVCTERIDETOLC  
AAPCLHRCIPCLTKMLPMRNSCPICNAVAVLVIVKDGSTRTPIVNDRTAE  
AEEAVRAGTVDTWTIRLPEGAAPSVTIGRTVRLSPKMGQAPRGAAPAP  
RAARVAPPPADSPILIAOSPAPSPRPVAGPPVAVPAPPRAPRAAPRAAP  
ALTOAQOTQARSAQAQAALQAALQALGRAPPRPQHAHTOTPRQAOTOTQ  
PTQARQOTQOTQOTQAOQAQAOTOTQOTQARPETOTOQOTQAOQAOTQOTQ  
RKPASGAGSAGSRGPKRASLPAPDAPARAPOLPRAPLAAAPRAPRAPASAS  
PRGSAPAPRAPRAPRAAGSSLGPRPARGRKCARKTHYDADRAPASGPTRYL  
PISGVSSVAMAPYLVNTKTVGDCLPVLDMEGTGAIGAVVIVGRCNLARCLADAEFQW  
ARRSRLEAPAPGCVSPPEYPGDPAHGLMTPVGMLFEQALGGRSFKSLDRHPWT  
PAEGDP"  
4856..4861  
complement(4879..4884)  
polya\_signal  
polya\_signal  
complement(8315..8320)  
9469..10167  
/gene="UL1"  
9469..10167  
/gene="UL1"  
/note="complexes with gH; involved in membrane fusion"  
/codon\_start=1  
/product="villon glycoprotein L"  
/protein\_id="AAU88066.1"  
/db\_xref="GI:52843275"  
/translation="MAVLRGVLLASLLFRCSFNNLLSTPADSSTTEYLIQSVAATV  
ADLVKVGICLDLPADGVTWRYEAPRSIDYARLIGLRYHQGIDTILMOGTARAAV  
NPLFTNGPFLIEDSHALFPNNALETARLTYKEVRLAUVASRGAASDTPVPGCTVS  
EYSTRKCLGRDTPGLFDEPRLRSLPUSAPGDEAPOPFSAPAPTSSAPPARRDRKPRP  
KPRNSTEPPRRR"  
10235..10990  
/gene="UL2"  
10235..10990  
/gene="UL2"  
10235..10990

polya\_signal  
polya\_signal  
gene  
CDS  
/gene="UL2"  
/note="proposed start codon is GTC"  
/codon\_start=1  
/transl\_except="(pos:10235..10237,aa:Met)  
/product="viral-DNA glycosylase"  
/protein\_id="AAU88067.1"  
/db\_xref="GI:52843276"  
/translation="MTFDSPARPALGDDTPRAGRPSPSLMPERRFRFLVGDAMPPLIE  
PELANPLTALNAEYERCRCAEVLPRFREDVSWTRCTPDEVAVVIGODYPHQEQ  
AHGLASVRRGTAPPSLNNVILTAARNCCBEATMSGHGLCKEAKRGVLLNTLTVR  
RGPAPHAQIGMDRFVGSVVRRLAASRPGILFVLMGMAHQAOSAIRPDPVIRVLTVAHP  
SPLSRVPFGSCRFCLANQYLRERSLAPIDWSR"  
11076..11081  
11130..11813  
/gene="UL3"  
11130..11813  
/gene="UL3"  
/note="colocalizes with ICP22 and UL4 in small, dense  
nuclear bodies"  
/codon\_start=1  
/product="UL3"  
/protein\_id="AAU88068.1"  
/db\_xref="GI:52843277"  
/translation="MSAVGAGAVPSTLAVASGWAFAAPHDGPAPARAPAPAPAP  
PTGHPVPPRADDRAPPGVDHVAFPMTMVSVDDELGRQLDITRKQRLSLV  
NLAICTKTSRSGTGRRARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR  
VIASRPRKRYTRSSDGRISPAPVVFHFVSDAPVHRDNYIAVPPRRRTSTILA  
DPASDPMQ"  
11809..11814  
complement(11832..11837)  
polya\_signal  
polya\_signal  
complement(11897..12511)  
/gene="UL4"  
complement(11897..12511)  
/gene="UL4"  
/note="colocalizes with ICP22 and UL3 in small, dense  
nuclear bodies"  
/codon\_start=1  
/product="UL4"  
/protein\_id="AAU88069.1"  
/db\_xref="GI:52843278"  
/translation="MSASASATIAVSLHGAATSSACVLPDAEQVCAFESEGTALAS  
RGCLRDAISRGAVVVRPTVGLVWVDCTREFCAVFGSRSEORLERHWDTGLCY  
PDSWSVSRGSAASPTAGIATVVGDSIYITVTVGSPGCEAGAPRRRLPPRPD  
APDAPSPVAVSPASATGAADLLEVMKEIQLSPTLGRPCDA"  
complement(12536..15163)  
/gene="UL5"  
complement(12536..15163)  
/gene="UL5"  
/codon\_start=1  
/product="component of helicase-primase complex"  
/protein\_id="AAU88070.1"  
/db\_xref="GI:52843279"  
/translation="MTTGESAGDASAPPPPEPPVAPARFDEAFINFTSMGIOP  
IRVYRELAGHRLPEAPRPPRPPWRDVALETPADLPLRPPPAVVLITNAGSGKST  
VTLASNEVDCCVVTGARITAAQNNYVLSAFLSRPINTIFHEGFGNHYQADGHP  
YALASPPASVEDQRRDLTYWEVIVADITRVAAGGGEDEGGERLAALERTGL  
AEGSLTRLAFTHGALPAPTRNSIVIDAGSLGRLLTAVVYCMWMTALVHTPOYA  
AGLRPVLVCGSPOTASLESTFEHQKSLROSENVTYILCNRTLEBYALPRSM  
AIFINNKRCVHEFGMLKALEYGLDTEHQSFVDRFVVPBSFITPNMLPGMTRLF  
SKHKEVSAVMARLHAHLKVAREGEFVFTLPVLTGTAFEREYRLTHOPLATVEKM  
IAANAGRITNYQSOQODGRRVRFVHSGQQLVVARNDVTYVLSGIAVTTTRKLVF  
GFGGTRSFEAVALRDSFVKTQGBMLVERAIFRSLILFGGLHFFNPLRGLDAK  
RAAVYERMGALTTLLALGAEAGAGGAAPAGARADFTLGAARAAGDAEPEDD  
LNVFAGLSEOEIDLFCYHVPEDPTTAVAHQFGLKRAPIGRYHILRELFGAEAF  
GAPFTYVNDNTFRCGEVWIGMRGLMSVALQOTDSYTLTGFTYAVVAFABVRERH  
AAAGLAELEESPLPVVLRDQHGFMVNTVSEVESVEABELAMATNADYGISSM  
LMTITRSGSLSDKVAICTPTGNLRANSAYVAMSTTSSERFLRMNLNLRERHEDD  
VLSGHTLSALRDEVLIV"  
15162..17198  
/gene="UL6"  
15162..17198  
/gene="UL6"  
/note="role in DNA cleavage/packaging"







TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Gabali, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hogue, M., Hollway, C., Hollins, B.,  
Homs, F., Howard, S., Huber, J., Huylk, S., Hume, J., Jackson, L.E.,  
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
Kratovic, J., Kureishi, A., Landry, R., Leal, B., Lewis, L.C., Lewis, L.,  
Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louieged, H.,  
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,  
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Moser, M., Neal, D., Newton, J., Newton, S., Ogih, M., Okumura, G.,  
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okumura, G.,  
Oragunye, N., Oyedro, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
Rivers, M., Rojas, A., Rojupokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
Scherer, S., Scott, G., Shen, H., Shooshitari, N., Stason, I.,  
Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansy, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Uemami, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G., and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 136753)

Worley, K.C.

Direct Submission

Submitted (25-JUN-2004) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: SPCA

Center clone name: R3-32A21

Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 128020 bases at least Q40

Consensus quality: 129713 bases at least Q30

Consensus quality: 130554 bases at least Q20

Estimated insert size: 129513; sum-of-coverage estimation

Quality coverage: 5x in Q20 bases; sum-of-coverage estimation

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 16 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of 'N', but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

1 1101: contig of 1101 bp in length  
\* 1102 1201: gap of unknown length  
\* 1202 2619: contig of 1418 bp in length  
\* 2620 2719: gap of unknown length  
\* 2720 4080: contig of 1361 bp in length  
\* 4081 4180: gap of unknown length  
\* 4181 5491: contig of 1311 bp in length  
\* 5492 5591: gap of unknown length  
\* 5592 6724: contig of 1133 bp in length  
\* 6725 6824: gap of unknown length

FEATURES

source

1. 136753  
/organism="Strongylocentrotus purpuratus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7668"  
/clone="R3-32A21"

ORIGIN

Query Match 12.8%; Score 124; DB 2; Length 136753;

Best Local Similarity 49.0%; Pred. No. 0.028;

Matches 398; Conservative 0; Mismatches 404; Indels 10; Gaps 3;

QY 31 GCGCGACGCGGCGGTGGCGCGCATGTCGCCGCGCTCTGCGGAGCGGCGACCGCG 90  
DB 3126 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3185  
QY 91 AGCGTGTGTGTGACCTCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 150  
DB 3186 GCG 3245  
QY 151 GTACACG 210  
DB 3246 GCG 3305  
QY 211 GCGGTACACGAGACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 270  
DB 3306 CCG 3365  
QY 271 GCG 330  
DB 3366 GCG 3424  
QY 331 CACG 390  
DB 3425 GCG 3484  
QY 391 ACCG 450  
DB 3485 GCG 3537  
QY 451 CGGCTGTGTGAGGCG 510  
DB 3538 GCG 3597  
QY 511 ACGACG 570  
DB 3598 GCGACG 3657  
QY 571 GCG 630  
DB 3658 CCG 3717







[illegible]

ACCESSION	AL684264
VERSION	AL684264.1
KEYWORDS	STS.
SOURCE	Penicillium marneffei
ORGANISM	Penicillium marneffei Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium
REFERENCE	1 Yuen,K.Y., Pascal,G., Wong,S.S., Glaser,P., Woo,P.C., Kunszt,F., Cal,J.J., Cheung,E.Y., Medicine,C. and Danchin,A. Exploring the Penicillium marneffei genome Arch. Microbiol. 179 (5), 339-353 (2003)
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	22595073
PUBMED	12640520
REFERENCE	2 (bases 1 to 1393) Danchin,A. and Pascal,G.
AUTHORS	Direct Submission
TITLE	Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre, Dexter Hc Man Building 8, Saseoon Road, Pokfulam, Hong Kong
JOURNAL	
FEATURES	Location/Qualifiers
source	1..1393 /organism="Penicillium marneffei" /mol_type="genomic DNA" /db_xref="taxon:37727" /clone="pm11h12.g"
ORIGIN	
Query Match	12.3%; Score 119.6; DB 11; Length 1393;
Best Local Similarity	45.6%; Pred. No. 0.45;
Matches	419; Conservative 0; Mismatches 485; Indels 14; Gaps 4;
Oy	60 CGCCGCGCTGTCTGGAGGCGGAGCGGCAGGCCGGAAGCCTGTCTGTCAACTGTACCGCGCC 119
Dd	1064 CCGCGGGGCG 1005
Oy	120 GGAGACGCCGCGAAGCCTTCCTTCTGTGTGGTGAGACTAACCGCGCGCGCGCGCGCGCGCG 179
Dd	1004 GCGGCG 945
Oy	180 CGCCCTCGCGGAGCGCGGACCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 239
Dd	944 GCGCCCG 885
Oy	240 CGTGCG 299
Dd	884 GCGGCG 825
Oy	300 CCGAAGCGGACATGAGCGCGCGGAGCTTCGCGGCGCGCACGCCCCCGCGCGTC---AGCATGTG 354
Dd	824 GCGGCG 765
Oy	355 GGCCTCAACCGAATGTTTCGCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 414
Dd	764 GNCGGCGAGC 705
Oy	415 GTCAACAAGAGCGAGCG 474
Dd	704 CCCG 645
Oy	475 AGGCGCGTAAGCTTCACGCGCGGAGAGCAACAACGGAAGAAGCGCGCGCGCGCGCGCGCG 533
Dd	644 GCGCGCGGCG 585
Oy	534 GAAGCAGCGCGCGTCTCTCTTCCTTCGAGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 593
Dd	584 GCTCG 525
Oy	594 GCGCTTGAGCG 653
Dd	524 CCG 465
Oy	654 GCTCGCGCGCGCGCG---CCGGAAGTGTACGCGGACATTCAGCGGTCAACCCCGCGCGCG 710



[illegible][illegible]



```
gene      complement (12172..14820)
/ gene= "UL5"
/ complement (12172..14820)
CDS       / gene= "UL5"
/ codon_start=1
/ product= "component of helicase-primase complex"
/ protein_id= "AAP41423.1"
/ db_xref= "GI:30844247"
/ translation= "MARVGDASVSSPPRPSPVOPRDEAFINFTSMHGIOTIER
VRLAHLRLAARVPRLOFRDVAALFTPGVLPURPPVYITLITAGSGKSTCVT
LNEVLCVVGATRIAQNNYVKLSGAFISRPINT1PHEFGFNNVQAOVGHPTL
ASNPASVEDLQRDLTYMEVIADITRAVAAGGADGDGNEFRALATLSTGLEG
ALTRLAFTGALPAFTRSNI VVI DEAGLGRHLTA VVYCMNTALYHTPOYAGL
RPLVYCGSPSTASLESTEHOKLRCSIRSENVIYLI CNLTRREVALPRSMALF
INMKCVHEHFGNLMKALEYGLPTEBHHOPURPVVPSFPTINPNALPGMTLPSH
KEYAYMARLHAHLKVARGEFVFSLPVLTIVSAGADEYRELHPALITVEKMLA
NAGRITNYSQSDQDAGRARLEVHSSKOQLVANRDVYVINSQ1AATTRLKLVFES
GTRSEPAVLRLDSFVKTQEMLVFEPYQFLSRILFGGLIFNFILRLPDAAARAR
AYERMGALITRELLSLGATGAPATPGQAASGGESASAAPGARAFDFTLGARAD
AEPEDDLVYFAGLNEOELDLFYCHYTVQPEPTTA VHAHQFLKRALFGRYHLIE
LFEAFEGAPGTYVNDNTPRGCEVMIGSRGLMVALOTDSYTLNGPYAVVAPA
BEYRRRAAGGLAELLEBSPLPYVLRDQGFMSVNTNISFVESSEHLMATNA
DYGISMLANTITRSQSLSDKVALICTPGNLRNLNSAYVAMSRTTSSEFLRNMLPLR
BRHERDVISSEHLSALRDPDLIVY"
14819..16870
/ gene= "UL6"
14819..16870
CDS       / gene= "UL6"
/ note= "role in DNA cleavage/packaging"
/ codon_start=1
/ product= "capsid protein"
/ protein_id= "AAP41424.1"
/ db_xref= "GI:30844248"
/ translation= "MAERHPRARPPARPPDAAAPRPDDGWDVQHPTTMTLREAL
GELGYTEGQGVNVVVRSSSEAVTQLOAAVFGALNAAHRRDLEADMRHVAARLRE
RLVRRYGAAPDGGVAAVAERKYDTPRAALASLLDFAGLDFAGGPGGPGFELRY
IDWMLCGLVPMVRPREGEVTORELSAFSLSDHLPDLATVAGAVRAAPGRLARA
FDSAR1ADYDRVCLFYHPRGGEWRVDRDPAAGRCGVLMIPMTGDRLLPDSMORL
SRRTVACRALREHARVCRILNNA SVYKLRKDGDRGAGARASRALGEDDASKA
GSAASRLVRLIIMKGMRHIGDINDTVRALDEAGHLIDSAVDALPGFGATGRP
AAGGARPAAPROQLQAFQTSVNNINLGEYINNLFGTIERLETYDIALTOQLD
RDIELRRARGALDROORAADPGARAEAPAGQDDRTVDIIDEVGAKDDTDVANS
FOHOVPSYAOGLERLSRLMEHELLRCFISRTNNOGRETAFVSGAIALPVAFY
ASVLRAPAMGALITGPNVYLGEELMEALFKTRVOTYTLIDLAALFVAVRRHAARF
PAPRPADPDRVARERPRGRSRSPVDAGRCGGRPEDDAMGGRGDHGLDPRGSDGGR
P"
16821..17711
/ gene= "UL7"
16821..17711
CDS       / gene= "UL7"
/ note= "role in DNA cleavage/packaging"
/ codon_start=1
/ product= "capsid protein"
/ protein_id= "AAP41425.1"
/ db_xref= "GI:30844249"
/ translation= "MATSTPAEBETAAALRRVADRDVGVVAEISHOALLRNAC
VRQATHTYRTATGVAVRVNVPDRHLHFLVLDCCPDALIASDAVFAKRDOSAHNGF
AFPAVILAREDDHNSLCPVLPVLRHSVPMEDLADPELACLIMLENDPRAATASA
PVKISAMLAALGRRTSPDRVRCILRS COMMLNTLMFWHVEFPDDQVYLPIMCAR
YLLANNPILTLFPAQPSRVFRLFPQPRASLCEVAVNPQGLVLAAGMSDEPRAGLV
SWMSSDAPKTOGSLFYRLV"
complement (17886..20171)
/ gene= "UL8"
/ complement (17886..20171)
CDS       / gene= "UL8"
/ codon_start=1
/ product= "component of helicase-primase complex"
/ protein_id= "AAP41426.1"
/ db_xref= "GI:30844250"
/ translation= "MGDAGLVMAEESICAITLYTAWLPRTDRDLVLLYLPCRDARG
DGAARPAEVSIASTELQAFYGAPEISAGAVAAAPAAAPALETLNDPVLMRALY
ACALVALEREVGRALFAPARVAMDPRTGIVARIEGLPGORSAPPAALDVNOAAT
DPLALAHHVAARPDARLAWARLAAIDTQGCASASALTVRIDGARGAFAREYTTTLFP
```

```

PVRRBGAADVFEVREAVLRRPRGHAQVATVAVLVPGRGYEFVAGEGFSAALIAMER
OMYLTVHAGRGALPPIFAFIAPGEFERBGAADVHDFSLGPRGMATLTKVSGAAAPPA
LAAHAEIAGAMPAAAGARALGPRARMAAVSAAAEERLPPALRRPAAAGLILDPAYIG
PVCLARFPGLQPIRLAALYEIGLGRDPRADRPDLGRGARMRDLGEBGAEOA
PMPREVLAATDEVRREYPAVARITLDGIVSAGVLEKAAADVNAVCDGGGRVWG
LVNDPRDAAGAAAAAERARALAAASSEALAGVGLDAPPLVLGRTYTHAVLVM

Query Match      12.1%; Score 117.4; DB 14; Length 156789;
Best Local Similarity 46.2%; Pred. No. 0.084;
Matches 391; Conservative 0; Mismatches 456; Indels 0; Gaps 0;

QY      6  CGGCTTCCCCCGGACGTCGTGTCGCGGCGACAGCGGGGCGGTGGCGGCAATGTTCCCGG 65
Db      153893  CGCGCGCGCGCGCGCGCTGTCGTCTGTGACCGCGGTGCCCGGAGAGATGCGCGGCCA 153952

QY      66  GCTGTCGGGAGAGCGGAGACGCCGACGCTGTCGACCTTCGTAACCGCCGCGGAGCG 125
Db      153953  CGGGCCCGCGTCGCGCGCGCGCCACGTCACCTGCGCTGCGCTGCGCGCGCGCGCA 154012

QY      126  GCGGAGCGCTGCTGTGGGGGAGAGTCACCGCGCGGGGGCCCGAATCGCGCGCGCCCT 185
Db      154013  GTGCGAGTCGCTGTGGCGCGCGCGCGGACCTGCGCGCGCGCTGTGCTGTGCGGCG 154072

QY      186  CGGGAGCGGAGACCTGTCCTGCTCGCTGACAGAGAGCTGGCCCTCAAGGCGGTGCG 245
Db      154073  CGTCTTGGGGCGCGCGCTGTCGCGCGCGCTGCGAGCGCGCTGCGCGCGCTCTACCCGGA 154132

QY      246  GCCCGTGAACCCGAGCTCATCGCGCGCGCGCGCTGTCGCGACCCCTGCTCGTGAGC 305
Db      154133  CGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 154192

QY      306  GGGGATGCGCGCGGAGACTCGCGGCGCCACGCGCGCGCGCTGCAAGCTGGGCTCAACC 365
Db      154193  CGGGCCCGACACGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTGCGCG 154252

QY      366  GATGTTGCGCGCGCGCGCGCGGATGACCGGCGGACCCGTCGCGCGCGCTGATCAAGGGA 425
Db      154253  GCTTCAGCGCGCGCGCGCGCGGACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 154312

QY      426  CGGGCGGGCGTCAACGCGCGCTGTCGAGGAGCGGCGGCGGAGCGCGCGCTAGC 485
Db      154313  CTTCCGCGCGCGGAGGCGGCGACCTCGACCGGCGCTGCGCGCGCTGCGGCGTGGCGCGCG 154372

QY      486  GCTCAGCGGAGAGACAGACCGGACGACGCGCGCTGACAGGCGCTTGAACGACGCCGT 545
Db      154373  GCTGCGGCGCGTCTACCTGCGCGCTGACCGGACCGCGCGCGCGCGCGCGCGCGCGCGCT 154432

QY      546  GCTCCTCTCTCTCGGCGCTGCGCGCGCGCGCGCTCGGCGCGCGCGCTCGGCGCGCG 605
Db      154433  GCGCGCGCGCGCTGCGCGCTTCTGCGCGCGCGCGCTGCTGAGCGGACCCCGACCGCG 154492

QY      606  GACGAGCACCGCGCGCCACAGAGTGTGCTGCTGCGCGCGCGCTGCTGCGCGCGCGAG 665
Db      154493  GCGCGCTCGTGTGCGCGCGGAGCGCGGAGCGCGGACCGCGCGCGCGCGCGCGGCTGCGTG 154552

QY      666  CCCCAGGTGTACGGGAGATCAAGCGGTCAACCCCGGAGCGCGCTCGCGCGCGCGCG 725
Db      154553  GGCCTGGGCGCGCGCGGAGACCTGCTCGCGGAGCGAGCGAGCGCGCGCGCGTGAAGGT 154612

QY      726  GCTCGCGGAGGCGCGCGCTGCTGCGCGCGCGCGCGTGGCGAGCGACCGGAGACCGTGC 785
Db      154613  GGCCTGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 154672

QY      786  CGCGCCCGGAGCGCGCGACGCGCGCGCGCGCATCGCGGAGATCGAGCGCGCGGAACT 845
Db      154673  CGCGCCCGGCGCGCGCGGAGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 154732

QY      846  CGACGCGC 852
Db      154733  CGCGCGC 154739

Search completed: October 5, 2005, 05:30:57
```



Job time : 4458.25 secs

---



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 16:46:09 ; Search time 585.401 Seconds  
(without alignments)  
9798.799 Million cell updates/sec

Title: US-10-089-514-5  
Perfect score: 969  
Sequence: 1 atgcgcgcgtccctcccgag.....aagcgagagaagacgatga 969

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1980s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES.

Result No.	Score	Query Match	Length	DB ID	Description
1	969	100.0	969	4 AAF86411	Aaf86411 4-amino-4
2	969	100.0	969	8 AAL50182	Aal50182 S venezue
3	915.4	94.5	3305	8 ABZ69799	Abz69799 Plasmid p
4	913.8	94.3	12391	8 ABZ69798	Abz69798 Plasmid p
5	270.8	27.9	888	2 AAT59269	Aat59269 Streptomy
6	270.8	27.9	2888	2 AAT59268	Aat59268 Streptomy
7	124.8	12.9	114955	4 AAX53491	Aax53491 Human ade
8	115.6	11.9	125401	4 AAD17186	Aad17186 Streptomy
9	115.4	11.9	114955	2 AAX53491	Aax53491 Human ade
10	114.8	11.8	65140	4 AAD17184	Aad17184 Streptomy
11	111.8	11.5	9222	8 AAL61171	Aal61171 Actinosyn
12	111.8	11.5	82746	8 AAL61224	Aal61224 Actinosyn
13	111.6	11.5	4770	8 AAL61172	Aal61172 Actinosyn
14	109.8	11.3	9975	8 AAL61173	Aal61173 Actinosyn
15	108.8	11.2	113193	8 AAD54645	Aad54645 Streptomy
16	107.4	11.1	969	4 AAF86411	Aaf86411 4-amino-4
17	107.4	11.1	969	8 AAL50182	Aal50182 S venezue
18	106.8	11.0	14055	8 AAL61170	Aal61170 Actinosyn
19	106.2	11.0	3305	8 ABZ69799	Abz69799 Plasmid p
20	105.4	10.9	12391	8 ABZ69798	Abz69798 Plasmid p

21	105.2	10.9	3354	8 AAL61181	Aal61181 Actinosyn
22	104.4	10.8	44377	2 AAT80414	Aat80414 Platenoli
23	104.4	10.8	44377	2 AAT80508	Aat80508 Platenoli
24	104	10.7	13842	3 AAZ87297	Aaz87297 S. venezu
25	104	10.7	13842	12 ADL91915	Adl91915 Streptomy
26	104	10.7	36778	3 AAZ87318	Aaz87318 S. venezu
27	104	10.7	36778	12 ADL91933	Adl91933 Streptomy
28	104	10.7	37948	3 AAZ87285	Aaz87285 S. venezu
29	104	10.7	38506	3 AAZ5633	Aaz5633 Nucleotid
30	104	10.7	38506	3 AAZ56001	Aaz56001 Recombina
31	104	10.7	38506	8 ADA09418	Ada09418 Cosmid pk
32	104	10.7	38506	10 ADH53462	Adh53462 S. venezu
33	104	10.7	38506	10 ABS56090	Abs56090 S. venezu
34	102.6	10.6	3957	5 AAO9606	Aao9606 HSV-2 imm
35	102.6	10.6	3957	10 ADG75121	Adg75121 Human her
36	102.6	10.6	154746	6 AAD25519	Aad25519 Human her
37	102.6	10.6	154746	6 AAD25519	Aad25519 Human her
38	101.4	10.5	5802	6 ABS78696	Abs78696 S. kaniba
39	101.4	10.5	15738	10 ADI23932	Adi23932 Streptomy
40	101.4	10.5	61944	10 ADI23920	Adi23920 Streptomy
41	101.4	10.5	109519	5 AAS08693	Aas08693 Micromono
42	100.4	10.4	2700	4 AAD17193	Aad17193 ERD48 ins
43	100.4	10.4	110000	4 AAI9682	Continuation (40 o
44	100.2	10.3	58857	3 AAS8471	Aas8471 Nucleotid
45	100	10.3	88421	6 AAL40781	Aal40781 88421nt g

## ALIGNMENTS

RESULT 1	AAF86411	standard; DNA; 969 BP.
ID	AAF86411	
XX	AC	AAF86411;
XX	DT	22-JUN-2001 (first entry)
XX	DE	4-amino-4-deoxyphenic acid dehydrogenase coding sequence.
XX	KM	Metabolite; benzene; chorismic acid; P-aminophenylpyruvic acid;
XX	OS	4-amino-4-deoxyphenic acid dehydrogenase; enzyme; pnpC; ds.
XX	XX	Streptomyces venezuelae.
XX	XX	Key
XX	XX	Location/Qualifiers
XX	XX	CDS
XX	XX	1..969
XX	XX	/product= "4-amino-4-deoxyphenic acid dehydrogenase"
XX	XX	WO200123542-A1.
XX	XX	05-APR-2001.
XX	XX	29-SEP-2000; 2000MO-JP006783.
XX	XX	29-SEP-1999; 99JP-00276314.
XX	XX	(MEIJ ) MEIJI SEIKA KAISHA LTD.
XX	XX	Yanaí K, Okakura K, Yasuda S, Watanabe M, Miyamoto K, Midoñ N;
XX	XX	Murakami T;
XX	XX	WPI; 2001-290517/30.
XX	XX	P-PDB; AAB82073.
XX	XX	Transformant producing secondary metabolite modified with functional
XX	XX	group e.g. benzene with nitrogen-containing substituent at para-position,
XX	XX	PT1022, with ease at low cost, for application in pharmaceuticals and
XX	XX	agrochemicals.
XX	XX	Claim 15; Page 67-70; 83pp; Japanese.



CC The present invention relates to transformant having been modified so  
CC as to produce a secondary metabolite. The secondary metabolite has a  
CC benzene ring skeleton free from substitution at the para-position by a  
CC nitrogen-containing functional group, thereby enabling the production of  
CC a secondary metabolite with a benzene ring skeleton substituted at the  
CC para-position by a nitrogen-containing group. The transformant organism  
CC of the present invention has been produced by transferring a gene  
CC participating in the biosynthesis pathway from *choismic acid* into *p*-  
CC *aminophenylpyruvic acid*. The present sequence is the coding sequence for  
CC 4-amino-4-deoxyisoprenic acid dehydrogenase (papC), from *Streptomyces*  
CC *venezuelae*. papC participates in the biosynthesis pathway from *choismic*  
CC *acid* into *p*-*aminophenylpyruvic acid*, and so the papC gene can be used to  
CC produce the transformant of the present invention. The transformant can  
CC be used to produce metabolites for application in pharmaceuticals,  
CC veterinary drugs and agrochemicals  
CC  
XX  
SO Sequence 969 BP; 106 A; 397 C; 358 G; 108 T; 0 U; 0 Other;

Query Match	100.0%	Score 969;	DB 4;	Length 969;
Best Local Similarity	100.0%	Pred. No. 2.8e-113;		
Matches 969;	Conservative 0;	Mismatches 13;	Indels 0;	Gaps 0;

OY	1	ATGACGAGCTTCCCCCGACGCTCGTCGCGGACGAGGAGGAGCGGTGTGGCCGACATGTTCC	60
Db	1	ATGAGCGGCTTCCCCCGACGCTCGTCGCGGAGGACGAGGAGCGGTGTGGCCGACATGTTCC	60
OY	61	GCCGAGCTGCTTCGGGAGGCGGGACAGCCGACGCTTCGTGCACTTCGTACCGCCGCGG	120
Db	61	GCCGAGCTGCTTCGGGAGGCGGGACAGCCGACGCTTCGTGCACTTCGTACCGCCGCGG	120
OY	121	GGAACGGCCGGAGACGCTTCGCTGTGGGCGACGTCACCGCGCGCGGGGCCGAACTCGCGGCC	180
Db	121	GGAACGGCCGGAGACGCTTCGCTGTGGGCGACGTCACCGCGCGCGGGGCCGAACTCGCGGCC	180
OY	181	GCCCTCCGAGGACGCGGACCTCGTCCCTGCGTCGCGCGTACAGAGGACGTGGCCCTCAAGGCC	240
Db	181	GCCCTCCGAGGACGCGGACCTCGTCCCTGCGTCGCGCGTACAGAGGACGTGGCCCTCAAGGCC	240
OY	241	GTCGCGCCCGGTGACCCCGCTCATGCGGCGCCGGCGCGCTCTCGCCACACCCCTGTCCGTC	300
Db	241	GTCGCGCCCGGTGACCCCGCTCATGCGGCGCCGGCGCGCTCTCGCCACACCCCTGTCCGTC	300
OY	301	CGGACGCGGACATGCGCGGAGCTTCGCGGCCCAAGGCCCGCGCGTTCAGACACGTGGGCTTC	360
Db	301	CGGACGCGGACATGCGCGGAGCTTCGCGGCCCAAGGCCCGCGCGTTCAGACACGTGGGCTTC	360
OY	361	AAACCCGATGTTCCGCCCGCGCGCGCGGACATGACCGCGCGACCCGTGGCCGCGGTGTACAC	420
Db	361	AAACCCGATGTTCCGCCCGCGCGCGCGGACATGACCCGTGGCCGCGGTGTACAC	420
OY	421	AGGACGAGGACGCGGCGGTCAACGAGCCCTGCTGCGGAGCTTCGTGAGAGGACGCGGACAGGCC	480
Db	421	AGGACGAGGACGCGGCGGTCAACGAGCCCTGCTGCGGAGCTTCGTGAGAGGACGCGGACAGGCC	480
OY	481	GTAACGGCTCAACGCGCGAGAGACGACGACCGGACGACGCGGCGGACCCAGGCCCTTGACGAC	540
Db	481	GTAACGGCTCAACGCGCGAGAGACGACGACCGGACGACGCGGCGGACCCAGGCCCTTGACGAC	540
OY	541	GCCGTGCTTCCTTCCTTGGGGCTCGGCCCTTCGCGCGCTTCGGGGGTGCAAGTCCGGGCTCTG	600
Db	541	GCCGTGCTTCCTTCCTTGGGGCTCGGCCCTTCGCGCGCTTCGGGGGTGCAAGTCCGGGCTCTG	600
OY	601	GCGGCGAGGACACCGCGCGCCACCAAGAGTGTGCTGCGCCCTCTGGCCCGGTGTGCTCGGCG	660
Db	601	GCGGCGAGGACACCGCGCGCCACCAAGAGTGTGCTGCGCCCTCTGGCCCGGTGTGCTCGGCG	660
OY	661	GCGAGACCCCGAGGTGTACGGGAGCAATCCAGCGGCTCAACCCCGGCGGCGGTCCGCGGCC	720
Db	661	GCGAGACCCCGAGGTGTACGGGAGCAATCCAGCGGCTCAACCCCGGCGGCGGTCCGCGGCC	720
OY	721	CGGCGCTTCGCGAGGCGCTTCCTTCGCGCGGCTGCTGTGAGCGACGACCCGACCGT	780
Db	721	CGGCGCTTCGCGAGGCGCTTCCTTCGCGCGGCTGCTGTGAGCGACGACCCGACCGT	780

[illegible]

RESULT 2	
AAL50182	
ID	AAL50182 standard; DNA; 969 BP.

AC AAL50182;

DT 28-JAN-2003 (first entry)

DE S venezuelae PF1022 substance gene #3

KW Substance PF1022; biosynthesis; chorismic acid; anthelmintic;

KW veterinary drug; gene; ds.

*Streptomyces venezuelae*.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

4.1	4.1: 202	a
4.1	/* tãa=	a
FT		

Product: F1072 substance

MOZ00211244-AL  
FN  
XX

FD 03-001-2002  
XX

PF 22-MAR-2002; 2002WO-JP002782.

PR 22-MAR-2001; 2001JP-00082227.

PA (MEIJ ) MEIJI SEIKA KAISHA LTD.

PI Yanai K, Sumida N, Watanabe M, Moriya T, Murakami T;

DR WPI; 2003-018934/01.

XX

PT substance derivatives by fermentation, as pharmaceuticals or veterinary

XX

[illegible]

CC PF1022 substance derivatives. These were obtained by transferring a gene

participating in the biosynthesis pathway from choribamic acid to p-aminophenylpyruvic acid (biosaithenase gene) into a phenylalanine-

CC The transFormants are producing PE1022 substance by  
CC requiring host derived from an organism producing the PE1022 substance.

CC fermentation, for use as pharmaceuticals or veterinary drugs. The present  
CC substance is a substance derived from Streptomyces

CC venezuela

SQ Sequence 969 BP; 106 A; 397 C; 358 G; 108 T; 0 U; 0 Other;

Query Match 100.0%; Score 969; DB 8; Length 969;







```
DB 2187 GGCCTCCGGAGACCGGACCTGCTGCTGCGGTACACGAGACGCGGCTCCCAAGGCC 2246
QY 241 GTGGCGCCCGGTGACCCGGGCTCATGCGGCGGGGCGCGCTGTGCGCGACACCTGTCGCTC 300
DB 2247 GTGGCGCCCGGTGACCCGGGCTCATGCGGCGGGGCGCGCTGTGCGCGACACCTGTCGCTC 2306
QY 301 CGGACGGGATGAGCCGCGAGCTCGCGGCCACGCCGCCGCTGCTCCAGACGATGGGCTTC 360
DB 2307 CGGACGGGATGAGCCGCGAGCTCGCGGCCACGCCGCCGCTGCTCCAGACGATGGGCTTC 2366
QY 361 AACCCGATGTTCCGCCCCCGCGGATGACCGCGGACCGCGTGGCCGCTGCTGCTACCC 420
DB 2367 AACCCGATGTTCCGCCCCCGCGGATGACCGCGGACCGCGGCTGCTGCTGCTGCTACCC 2426
QY 421 AGGGACGGGCGGGGCTGACCGGCTGCTGCGGCTGCTGCGAGGGCGGGCGGCGGCGGCGCC 480
DB 2427 AGGGACGGGCGGGGCTGACCGGCTGCTGCGGCTGCTGCGAGGGCGGGCGGCGGCGGCGCC 2486
QY 481 GTACGGCTCAACGCGGAGAGACGACCGGACGACCGGCGGCGACCGGCGCTGACGCGAC 540
DB 2487 GTACGGCTCAACGCGGAGAGACGACCGGACGACCGGCGGCGACCGGCGCTGACGCGAC 2546
QY 541 GCGGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
DB 2547 GCGGTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2606
QY 601 GCGGCGACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
DB 2607 GCGGCGACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2666
QY 661 GCGGCGGCGGAGGTGATCGGAGGACATCGACGGGTGCAACCGCGGCGGCGGCGGCGGCGG 720
DB 2667 GCGGCGGCGGAGGTGATCGGAGGACATCGACGGGTGCAACCGCGGCGGCGGCGGCGGCGG 2726
QY 721 CGGGGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 768
DB 2727 CGGGGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2786
QY 769 -----GACCGGAGCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 822
DB 2787 GCGGAGGACCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2846
QY 823 GATGCGACGCGCGCGGAGAACTCGACGCGCTTTCGCGGAACTTCGCGCGGCTCATGAGG 882
DB 2847 GATGCGACGCGCGCGGAGAACTCGACGCGCTTTCGCGGAACTTCGCGCGGCTCATGAGG 2906
QY 883 CCGGAGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 942
DB 2907 CCGGAGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2966
QY 943 GACGACGAGGCGGAGGAGGACGCGATGA 969
DB 2967 GACGACGAGGCGGAGGAGGAGGACGATGA 2993

RESULT 4
ABZ69798
ID ABZ69798 standard; DNA; 12391 BP.
XX
AC ABZ69798;
XX
DT 08-APR-2003 (first entry)
XX
DE Plasmid plasc-papabc.
XX
XX DHFR; translation; orthogonal tRNA; O-tRNA; O-RS; CAT; TyrRS;
KW orthogonal aminoacyl tRNA synthetase; unnatural amino acid;
KM chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;
XX biosynthesis; p-aminophenylalanine; paf.
XX
XX Synthetic.
```

```
PN W0200285923-A2.
XX
XX 31-OCT-2002.
XX
PF 19-APR-2002; 2002W0-US012465.
XX
PR 19-APR-2001; 2001US-0285030P.
XX
PR 06-FEB-2002; 2002US-0355514P.
XX
PA (SCRI ) SCRIpps RES INST.
XX
PI Schultz P, Wang L, Anderson JC, Chin JWK, Liu DR, Magliery TU;
PI Meggers EL, Mehl RA, Pasternak M, Santoro SW, Zhang Z;
XX
XX MPI; 2003-120430/11.
PT Composition useful for producing protein comprising unnatural amino acid,
PT has translation system comprising orthogonal tRNA and orthogonal
PT aminoacyl tRNA synthetase.
XX
XX Example 4; Page 124-127; 188pp; English.
XX
XX CC The invention relates to a novel composition comprising a translation
XX CC system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl
XX CC tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O
XX CC -tRNA with at least one unnatural amino acid in the translation system
XX CC and the O-tRNA recognises at least one selector codon. A composition of
XX CC the invention is useful for producing at least one protein comprising at
XX CC least one unnatural amino acid. The protein is the Asp127NG mutant of
XX CC chloramphenicol acetyltransferase (CAT), the Tyr1637NG mutant of mouse
XX CC dihydrofolate reductase (DHFR), or a Tyr1637NG mutant of mouse
XX CC dihydrofolate reductase comprising CHO0H His6tag. The unnatural amino
XX CC acid is provided exogenously. The translation system is a cell and the
XX CC unnatural amino acid is biosynthesised by the cell. The present sequence
XX CC represents a plasmid for use in the biosynthesis of p-aminophenylalanine
XX CC (PAF) in vivo
XX
SQ Sequence 12391 BP; 2830 A; 3588 C; 3346 G; 2627 T; 0 U; 0 Other;
XX
XX Query Match 94.3%; Score 913.8; DB 8; Length 12391;
XX Best Local Similarity 96.5%; Pred. No. 1.2e-106;
XX Matches 952; Conservative 0; Mismatches 17; Indels 18; Gaps 1;
QY 1 ATGAGCGGCTTCCCGCGACGCTGCTGCTCGGCGGAGCGGAGCGGTGGGCGGCGATGTC 60
DB 2361 ATGAGTGGCTTCCCGCGAGCGCTGCTGCTCGGCGGAGCGGAGCGGTGGGCGGCGATGTC 2420
QY 61 GCGGCGCTGCTGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
DB 2421 GCGGCGCTGCTGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2480
QY 121 GGAAGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
DB 2481 GGAAGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2540
QY 181 GCGCTTCGGGAGCGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
DB 2541 GCGCTTCGGGAGCGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2600
QY 241 GTGGCGCCCGGTGACCCGGGCTCATGCGGCGGGCGGCGGCTGCTGCGACACCTGTCGCTC 300
DB 2601 GTGGCGCCCGGTGACCCGGGCTCATGCGGCGGGCGGCGGCTGCTGCGACACCTGTCGCTC 2660
QY 301 CGGACGGGATGAGCCGCGAGCTCGCGGCCACGCCGCCGCTGCTCCAGACGATGGGCTTC 360
DB 2661 CGGACGGGATGAGCCGCGAGCTCGCGGCCACGCCGCCGCTGCTCCAGACGATGGGCTTC 2720
QY 361 AACCCGATGTTCCGCCCCCGCGGATGACCGGCGGACCGCGGCTGCTGCTGCTGCTGCTGCTG 420
DB 2721 AACCCGATGTTCCGCCCCCGCGGATGACCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTG 2780
QY 421 AGGGACGGGCGGGGCTGACCGGCTGCTGCGGCTGCTGCGAGGGCGGCGGCGGCGGCGGCGCC 480
```



Dd	2781	AGGACGGGGCGGGGGCTCA	CAGGCCCTGTGTCGGGCTCTTTCAGAGGGCGGGGGGCAGAGGCC	2841										
Oy	481	GTA	CGGCTCA	CGGCGGAGAGCAGCACCGGACGACGCGCGCCACCAGGCCCTTGACGAC	540									
Dd	2841	GTAC	CGGCTCA	CGGCGGAGAGAGCAGCAGCCGGAAGCAGCGCGGAGAACCGGCCCTTGACGAC	2900									
Oy	541	GCCG	GGCTCTCTCTTTGGGGTGGCCCTCCGCGCCCTGGCGTTCGAGTCCGGGCCCTG	600										
Dd	2901	GCCG	GATCTCTCTCTTGGGGCTCGCCCTCCGCGCCCTGGCGTTCGAGTCCGGGCCCTG	2966										
Oy	601	GCGG	CGGACGGGAC	CGGCGCCGCCACCAAGGTGCTCGCCCTTCCTGGCCGCTGGCTCGGC	660									
Dd	2961	GCGG	CGGACGGGAC	CGGCGCCGCCACCAAGGTGCTCGCCCTTCCTGGCCGCTGGCTCGGC	3020									
Oy	661	GCGAC	CCCCCGAGTGTACG	GGGACATCCAGCGGTCMAA	CCCGCGGCGCGTCCGCGGC	720								
Dd	3021	GGGAC	CCCCCGAGTGTACG	GGGACATCCAGCGGTCMAA	CCCGCGGCGCGGTCGCGGC	3080								
Oy	721	CGGG	GCTTCGCGAGG	CCCCTTGCGCTCTTCCGCGGCTGGTGGGAC-----	768									
Dd	3081	CGGG	GCTTCGCGAGG	CCCCTTGCGCTCTTCCGCGGCTGGTGGGAC-----	768									
Oy	768	-----	GA	CCCCGGACCGTGGCGGACGCCCCCGGGGGCGCGGACGCCCCCGGCGCATCCGGG	822									
Dd	3141	GCCG	GAGAC	CGCGGACCGCGGACCGCCGACCGCCGACCGACGACAA	CCCGGCGCATCCGGG	3200								
Oy	823	GGAT	GCGAC	CGGCGCGCGGGA	CCTCGACGCGCTCTTCCGGGGA	CTCCGCGCTCATGGGA	882							
Dd	3201	GGAT	GCGAC	CGGCGCGCGGGA	CCTCGACGCGCTCTTCCGGGGA	CTCCGCGCTCATGGGA	3266							
Oy	883	CCG	GAGCTTCGCGG	CGGCGCGGACCACTG	CCAGAGCTGTTCCGAC	CCCTTCA	CGGAC	942						
Dd	3261	CCG	GAGCTTCGCGG	CGGCGCGGACCACTG	CCAGAGCTGTTCCGAC	CCCTTCA	CGGAC	3320						
Oy	943	GACG	ACGAA	GGCAGA	GAAGACCGATGA	969								
Dd	3321	GACG	ACGAA	GGCAGA	GAAGACCGATGA	3347								
<hr/>														
RESULT 5														
AAT59269 standard; cDNA; 888 BP.														
XX AC	AAT59269;													
XX DT	02-APR-1997	(first entry)												
XX DE	Streptomyces pristinaespiralis papC gene.													
XX XX	Streptogramin B; antibiotic; biosynthesis; pristinamycin; virginiamycin;													
KW KM	DMPAPA precursor; 4-dimethylamino-L-phenylalanine; pApM; pApB;													
KW KW	pApC; isomerisation; aromatisation; N-methyltransferase; ds.													
XX OS	Streptomyces pristinaespiralis.													
XX FH	Key	Location/Qualifiers												
FH FT	mat_peptide	1..888												
FT FT		/tag= a												
FT FT		/product= "PapC"												
PN PN	WO9601901-A1.													
XX PD	25-JAN-1996.													
XX PF	04-JUL-1995;	95MO-FR000889.												
XX PR	08-JUL-1994;	94FR-0008478.												
XX PA	(RHON ) RHONE POULENC RORER SA.													
XX PI	Blanc V, Thibaut D, Bamas-Jacques N, Blanche F, Crouzet J,													
XX PT	Barriere J, Debussche L, Famechon A, Paris J, Dutruc-Rosset G;													
XX XX														

[illegible]











AD	AAD17186	standard; DNA; 125401 BP.
AC	AAD17186;	
DT	29-NOV-2001	(first entry)
DE	Streptomyces noursei	nystatin PKS gene cluster DNA.
KW	Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster; antifungal; antibiotic; ds.	
OS	Streptomyces noursei.	
FH	Key	Location/Qualifiers
FT	CDS	6337..34771
FT		/tag= a
FT	CDS	/product= "NysI complete protein"
FT		34792..51099
FT		/tag= b
FT	CDS	/product= "NysJ protein"
FT		51155..57355
FT		/tag= c
FT	CDS	/product= "NysK protein"
FT		57503..58687
FT		/tag= d
FT	CDS	/product= "NysL protein"
FT		complement(58786..58980)
FT		/tag= e
FT	CDS	/product= "NysM protein"
FT		/note= "CDS does not include start codon"
FT		complement(59045..60241)
FT		/tag= f
FT	CDS	/product= "NysN protein"
FT		/note= "CDS does not include start codon"
FT		complement(60238..61296)
FT	CDS	/tag= g
FT		/product= "NysO complete protein"
FT		120628..121308
FT		/tag= h
FT		/product= "NysR (long) protein"
XX	MO200159126-A2.	
XX	16-AUG-2001.	
PD		
PF	08-FEB-2001; 2001WO-GB000509.	
PR	08-FEB-2000; 2000GB-00002840.	
PR	10-APR-2000; 2000GB-00008786.	
PR	14-APR-2000; 2000GB-00009387.	
PA	(UNO-) UNIV NORGE TEKNIKSK NATURVITENSKAPELIGE.	
PA	(SWF ) SINTEF STIFTELSEN IND TEK FORSK.	
PA	(ALPH-) ALPHARMA AS.	
PA	(SINV-) SINVENT AS.	
PA	(DZIE/) DZIELEWSKA H.	
PA	(ZOTC/) ZOTCHEV S B.	
PA	(SEKU/) SEKUROVA O N.	
PA	(FOAE/) FOAERVIK E.	
PA	(BRAU/) BRAUTASET T.	
PA	(STRO/) STROM A R.	
PA	(VALL/) VALLA S.	
XX		
EI	Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR,	
FI	Valla S, Ellingsen TE, Sletta H, Gulliksen O;	
XX		
DR	WP1: 2001-557614/62.	
DR	P-PSDB: AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,	
DR	AAE10149, AAE10150.	
XT	New nystatin polyketide synthase polynucleotides and polypeptides, useful as antibiotics and antifungals.	
XT		

[illegible]



Db 229394 ACGCCTTCGCCGGCGCTCGCGGACACCGGCGACACCGGACACTGGCGCTGCTGGGGGG 29453

Oy 908 ACTGCCAGAGACTGTTCCGCACCTCCACCGCAGCCAGC 945

Db 29454 CCTGGCCCGACCTGCTGTCGCGCCGCTGCTCCGCGGCAC 29491

RESULT 9

AAK53491

XX AAK53491 standard; DNA; 114955 BP.

XX

XX AAK53491;

XX

XX 05-JUL-1999 (first entry)

XX Human adenosine A1 receptor antisense oligonucleotide fragment.

XX

XX Antisense oligonucleotide; multiple target; antisense treatment;

XX impaired respiration; inflammation; lung disease;

XX pulmonary vasoconstriction; inflammation; allergic rhinitis;

XX acute asthma; allergy; asthma; impeded respiration;

XX respiratory distress syndrome; pain; cystic fibrosis;

XX pulmonary hypertension; pulmonary vasoconstriction; emphysema;

XX chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;

XX colon cancer; breast cancer; lung cancer; pancreatic cancer;

XX hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;

XX prostate cancer; ss.

XX

XX Synthetic.

XX

XX WO9913886-A1.

XX

XX 25-MAR-1999.

XX

XX 17-SEP-1998; 98WO-US019419.

XX

XX 17-SEP-1997; 97US-0059160P.

XX

XX 09-JUN-1998; 98US-00093972.

XX

XX (UYEC-) UNIV EAST CAROLINA.

XX

XX MYCE JW;

XX

XX WPI; 1999-229400/19.

XX

XX New antisense oligonucleotides used in treatment of, e.g. pulmonary

XX vasoconstriction.

XX

XX Disclosure; Page 37; 120BP; English.

XX

XX The specification describes antisense oligonucleotides (AAK52869-X55271)

XX directed against at least 2 mRNAs selected from target genes, coding and

XX non-coding regions of RNAs corresponding to target genes, gene initiation

XX codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'-

XX end and the junction-section between coding and non-coding regions and all

XX segments of RNAs encoding proteins associated with one or more diseases,

XX conditions or mixtures. The antisense oligonucleotides may be derived

XX from sequences AAK55272-74. These multiple target oligonucleotides

XX (specifically AAK55180-271) can be used for the antisense treatment of

XX diseases and conditions. Typical diseases and conditions are those

XX associated with impaired respiration and inflammation, including lung

XX diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis,

XX acute asthma, allergies, asthma, impeded respiration, respiratory

XX distress syndrome, pain, cystic fibrosis, pulmonary hypertension,

XX pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary

XX disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g.

XX colon cancer, breast cancer, lung cancer, pancreatic cancer,

XX hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as

XX well as all types of cancers which may metastasize or have metastasized

XX to the lungs, including breast and prostate cancer

XX

XX Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 0 U; 21427 Other;

Query Match	11.9%; Score 115.4; DB 2; Length 11495;
Best Local Similarity	34.2%; Pred. No. 6.4e-07;
Matches	321; Conservative 101; Mismatches 502; Indels 15; Gaps 3;
OY	15 CCGCAGCGCTGCTGTCCGGCGCAGCGGGGCGGTGGGCGGCATGTTGCCGCGCTGTGCG 74
Db	1039936 CGGCGTTCGGGCGCGGGCGSNNNNDNNTTGGCGGCTGTCCGGGCSNNNDNNGCGGCGTCCG 1040555
OY	75 GGAGGCGGAGCAGCGCACGCTGTGTGTGCACTCTGTACCGCCCGGAGCGCGGACGC 134
Db	104056 GCCGGGCSNNNDNNCGGGCGTGCGGCGCGGCSNNNDNNGCGGTGTGGCGCGGCSNNDN 104115
OY	135 CTGCTGCTGTGGGCGACGTACCGCGCGCGGGGCGCCGAATCTCGCGCGCCCTCTCGGAGCG 194
Db	104116 NCGCGTCCGGCCGGGCSNNNDNNGCGTGTGGGCGGGGCSNNNDNCTGTGGGCGGGCSNN 104175
OY	195 GGACTCTGCTCTGTGCGCGTACAGAGGACGTGGCCCTCAAGGCGCGTGGCGCGCTGAC 254
Db	104176 NDNNGTCCGGGCGGGCSNNNDNNTTGGGCGGGGCSNNNDNCGGCGCGGCSNNNDNCC 104235
OY	255 CCGGCTCATGTCCGC-----CGGCGCGCTGTCTGCGGACACCCTGTCTCGTCCGAGCG 306
Db	104236 GCBGCGCBGGGCGCGCGCGCGCGCGGCGSNNNDNCCGCBGGCGCBGGGCGCGCGCG 104295
OY	307 GGCATGCGCGCGGAGCTGCGGGGCCACGCCCCCGCGTCAAGACGTGGGCTCAACCG 366
Db	104296 GCGGCGCSNNNDNCCGCBGGCGCBGGCGCGCCCGCGGCSNNNDNCCGCBGGC 104355
OY	367 ATGTTGCGCCCGCGCGCGGACGTACCGGCGCACCCGTGTGGCGCGGTGTCAACAGGAC 426
Db	10436 CBGGGCGCGCGCGCGCGCGGGSNNNDNCCGCBGGCBGGGCGCGCGCGCGCGSSNN 104415
OY	427 GGGCGGCGGCTGTCAGCG-----CCTGTGTCGCGCTGTGTGAGGCGGCGGCGAGGCC 480
Db	104416 DNNCCGCBGGCGCBGGGCGCGCGCGCGCGGGSNNNDNCCGCBGGCGCBGGGCGCGCGCG 104475
OY	481 GTACGCTCACGCGGAGAGACGACCGAGAGGACGCGCGGCCAACCCAGGCTGTGAC 540
Db	104476 GCCSNNNDNCCGCBGGCGCBGGGCGCGCGCGSGSNNNDNCCGCBGGCBGGGCGCGCG 104535
OY	541 GCCGTGCTCTCTCTTCTTGGGGCTGCGCCTCGCGCGCTGTGGGTGACGTCCGGGCGCTG 600
Db	104536 CGCCGGSNNNDNCCGCBGGCGCBGGGCGCGCGCGSGSNNNDNCCGCBGGCBGGGCGCG 104595
OY	601 GCGGCGAGCGGACCGCGCGCCCACAGGAGTGTGTCTCGCCCTCTGGGCGGTGTGCTCGGC 660
Db	104596 CCGCSNNNDNCCGCBGGCGCBGGGCGCG--GCGCSNNNDNCCGCBGGCGCBGGGCGCGCG 104654
OY	661 GGCAGCCCCGAGGTGTACGGGAGCATCTCAGCGGTCAAACCCCGGCGCGCTCGCGCGC 720
Db	104655 GSNNNDNCCGCBGGCGCBGGGCGCGCSNNNDNCCGCBGGCBGGGCGCGCSNNNDNCC 104714
OY	721 CGGGGCTGTGCGGAGGCGCTGTCTTGTGCGCGCTGTGTGTGGGAGCGAACCAGACGT 780
Db	104715 GCBGCGCBGGGCGCGGSNNNDNCCGCBGGCGCBGGGCGCSNNNDNCCGCBGGCBGGGCG 104774
OY	781 GCCGACGCGCCCGCGCGCGCGGACGCCCCCGGCGCAATCCCGGGAGATGCGACGCGCGCGG 840
Db	104775 GSNNNDNCCGCBGGCGCBGGGCSNNNDNCCGCBGGCBGGGCSNNNDNCCGCBGGCBGG 104834
OY	841 AACCTGACGCGGCTTTCGGGGAATCTCCGCGGCTCATGAGACCGGAGCTCGCGCGGCG 900
Db	104835 GSNNNDNCCGCBGGCGCBGGGCSNNNDNCCGCBGGCBSNNNDNCCGCBGGCGSSNNNDNCC 104894
OY	901 CAGGACCACTGCCAGAGACTGTTCGCAACCTCCACCGC 939
Db	104895 GCBGGCGBGGGCGCGCGCGCGCGCGCGSGSNNNDNCC 104933
RESULT 10	
ADD17184	
ID	ADD17184 standard; DNA; 65140 BP.
XX	















QY 126 GCCGAGCCTGCTGCTGAGCGACGTCACCGCCGAGGCGCGACTCGGCGGCT 185  
 DB 2790 GCTGCGGCGGCTGCTGCGCCGCGGCTGACCGCTGCGACAGAGCGAGCCAGCGGCGCT 2849  
 QY 186 CCGGAGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 245  
 DB 2850 GCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2906  
 QY 246 GCCGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 305  
 DB 2907 CCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2966  
 QY 306 GCGCATGAGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 365  
 DB 2967 GCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3026  
 QY 366 GATGTTGCGCCCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 425  
 DB 3027 CCGCGACACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3086  
 QY 426 CCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 485  
 DB 3087 CAGGCGCTGCGC--GACGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3143  
 QY 486 GCTCACGCGCGGAGGAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 545  
 DB 3144 GCTGCGATCTCCCGCGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 3203  
 QY 546 GCTCTCTCTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 605  
 DB 3204 GCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3263  
 QY 606 GAGGAGCAGCT 665  
 DB 3264 CCGCT 3323  
 QY 666 CCGCGAGGCTGACGCGGAGCAGTCAGCGCGCTCAACCGCGCGCGCGCGCGCGCGCGCT 725  
 DB 3324 CAGAGGAGGAGGAGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3383  
 QY 726 GCTGCGGAGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 785  
 DB 3384 GCGCGCGGAGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCT 3443  
 QY 786 CCGCT 845  
 DB 3444 CCGCGGAGCAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3503  
 QY 846 CAGCGCGCTCTTGGGAGACGTCGCGCGCGCTCATGAGACCGGAGCTGCGCGCGCGCGAG 905  
 DB 3504 CAGCTGCTGCTCTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCT 3563  
 QY 906 CCACTGCGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 955  
 DB 3564 CAGCTGAGAGCGCTGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3613

RESULT 14  
 AAL61173  
 ID AAL61173 standard; DNA; 9975 BP.  
 XX AAL61173;  
 AC  
 DT 22-SEP-2003 (first entry)  
 XX Actinosynema pretiosum polyketide synthase (PKS) gene #4.  
 DB Maytansinoid; ansamitocin; antitumour; polyketide synthase; enzyme; PKS;  
 KM gene; ds.  
 XX Actinosynema pretiosum.  
 OS  
 PN WO2003045312-A2.

XX 05-JUN-2003.  
 PD 21-NOV-2002; 2002WO-US037547.  
 PF 21-NOV-2001; 2001US-0332158P.  
 PR 21-NOV-2001; 2001US-0332158P.  
 XX  
 PA (UNIW ) UNIV WASHINGTON.  
 XX Floss HG, Yu T, Leistner E;  
 PI  
 XX  
 DR WPI; 2003-493374/46.  
 XX  
 PT Novel maytansinoid produced by bacterial host cell transformed with  
 PT expression vector comprising open reading frame from ansamitocin gene  
 PT cluster I of Actinosynema pretiosum, useful as antitumor agent.  
 XX  
 PS Disclosure; Page 62-67; 160pp; English.  
 CC  
 CC The invention relates to maytansinoid produced by bacterial host cell  
 CC transformed with expression vector comprising open reading frame from  
 CC ansamitocin gene cluster I of Actinosynema pretiosum. Maytansinoid is  
 CC useful as a potent antitumour agent. The present sequence is A. pretiosum  
 CC ansamitocin gene cluster I polyketide synthase (PKS) gene  
 CC  
 SQ Sequence 9975 BP; 947 A; 4102 C; 3795 G; 1131 T; 0 U; 0 Other;  
 Query Match 11.3%; Score 109.8; DB 8; Length 9975;  
 Best Local Similarity 46.8%; Pred. No. 5,9e-06;  
 Matches 414; Conservative 0; Mismatches 467; Indels 4; Gaps 2;  
 QY 69 GCTGCGGAGGCGGAGCT 128  
 DB 3716 GCGGCGAGGCT 3775  
 QY 129 GAGCGCTGCGC-TGCTGAGGAGCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCT 187  
 DB 3776 CCGCGCGAGGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 3835  
 QY 188 GAGAGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 247  
 DB 3836 ACGACTGATCACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3895  
 QY 248 CCGTACCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 307  
 DB 3896 GGGTCAACCGCGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3955  
 QY 308 GCATGCGCGGAGCTGCT 367  
 DB 3956 CCGAGGAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4015  
 QY 368 TGTTCGCT 427  
 DB 4016 ACCTGTCACCGCT 4075  
 QY 428 GCGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 487  
 DB 4076 CCGCTGCT 4135  
 QY 488 TCAGCGGAGGAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 544  
 DB 4136 GCGCGCTGTCAGCTGCT 4195  
 QY 545 TGCTCTCTCTTTCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 604  
 DB 4196 CCGACCGCTCACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4255  
 QY 605 CAGAGGAGCT 664  
 DB 4256 CGCTCACCGCT 4315  
 QY 665 GCCCGAGGAGTACGCGGAGCATCAGCGGCTCAACCGCGCGCGCGCGCGCGCGCGCT 724



```
Db      4316 TGCAGAGCTGACCTCCGACCTGACGCTGTTTCGCTCTCTCCGACCTCCGCGATC 4375
Qy      725 CGCTCGCGGAGGCGCTCGCTCTTCGCGCGCGCGTGGCGAGACCGGACCGTGGCG 784
Db      4376 TCGGCGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4435
Qy      785 AGCGCGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 844
Db      4436 GCGAGCGCGACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4495
Qy      845 TCGACGCGCGCTTCGCGGGAACCTCCGCGGCTCATGGAGCGGAGCTCGCGCGGCGGCGG 904
Db      4496 CGAGCGGCGTACCGGCTCCCTGTCGCGGACCGACCGGAGCGGATGGCCGCTCCGCGG 4555
Qy      905 ACCACTGCGCAGAGCTGTTCCGACCGCTCCACCGGACCGGACG 949
Db      4556 TCGCGCGCGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4600

RESULT 15
AADS4645 ID AADS4645 standard; DNA; 113193 BP.
XX
AC AADS4645;
XX
DT 26-JUN-2003 (first entry)
XX
DE Streptomycetes nodosus amphotericin (amph) biosynthetic gene cluster.
XX
KM Polyene; antibiotic; amphotericin; amph; polyketide; enzyme; gene; ds.
XX
OS Streptomycetes nodosus.
XX

Key      Location/Qualifiers
CDS      FH      /tag= a
          FT      /product= "ABC transporter encoded by S. nodosus amphG
          FT      gene"
          FT      complement(1805..3628)
          FT      /tag= b
          FT      /product= "ABC transporter encoded by S. nodosus amphH
          FT      gene"
          FT      3840..4874
          FT      /tag= c
          FT      /product= "GDP-mannose dehydratase encoded by S. nodosus
          FT      amphiIII gene"
          FT      5042..33574
          FT      /tag= d
          FT      /product= "Polyketide synthase multienzyme housing
          FT      extension modules 9, 10, 11, 12, 13 and 14 encoded by S.
          FT      nodosus amphi gene"
          FT      33584..50518
          FT      /tag= e
          FT      /product= "Polyketide synthase multienzyme housing
          FT      extension modules 15, 16 and 17 encoded by S. nodosus
          FT      amphiJ gene"
          FT      50571..56675
          FT      /tag= f
          FT      /product= "Polyketide synthase multienzyme housing
          FT      extension modules 18 and thioesterase encoded by S.
          FT      nodosus amphi gene"
          FT      56829..58019
          FT      /tag= g
          FT      /product= "Cytochrome P450 encoded by S. nodosus amphi
          FT      gene"
          FT      58139..58648
          FT      /tag= h
          FT      /product= "ORF1, hypothetical protein"
          FT      complement(58756..59610)
          FT      /tag= i
          FT      /product= "ORF2, hypothetical protein"
          FT      59869..61470
          FT      /tag= j
          CDS
```

```
FT      /product= "ORF3, hypothetical protein"
FT      /transl_except= (pos:59869..59871, aa:Met)
FT      complement(61798..61995)
FT      /tag= k
FT      /product= "Ferredoxin encoded by S. nodosus amphiM gene"
FT      complement(62051..63250)
FT      /tag= l
FT      /product= "Cytochrome P450 encoded by S. nodosus amphiN
FT      gene"
FT      complement(63250..64308)
FT      /tag= m
FT      /product= "NDP-sugar aminotransferase encoded by S.
FT      nodosus amphiDII gene"
FT      complement(64324..65775)
FT      /tag= n
FT      /product= "Glycosyl transferase encoded by S. nodosus
FT      amphiI gene"
FT      /transl_except= (pos:65773..65775, aa:Met)
FT      66081..70319
FT      /tag= o
FT      /product= "Polyketide synthase multienzyme housing
FT      loading module encoded by S. nodosus amphiA gene"
FT      70366..79938
FT      /tag= p
FT      /product= "Polyketide synthase multienzyme housing
FT      extension modules 1 and 2 encoded by S. nodosus by amphiB
FT      gene"
FT      79956..112709
FT      /tag= q
FT      /product= "Polyketide synthase multienzyme housing
FT      extension modules 3, 4, 5, 6, 7 and 8 encoded by S.
FT      nodosus by amphiC gene"
XX
PN WO200297082-A2.
XX
PD 05-DEC-2002.
XX
XX 27-MAY-2002; 2002WO-1E000071.
XX
PF 31-MAY-2001; 2001IE-00000527.
XX
PR (UYDU-) UNIV COLLEGE DUBLIN.
XX
PA Caffrey JP;
XX
PI WPI: 2003-201271/19.
XX
DR P-PSDB; AAE36116, AAE36117, AAE36118, AAE36119, AAE36120, AAE36121,
XX P-PSDB; AAE36122, AAE36123, AAE36124, AAE36125, AAE36126, AAE36127, AAE36128,
DR AAE36129, AAE36130, AAE36131, AAE36132.
XX
PT Novel cytochrome P450 enzyme and nucleotides encoding the enzyme, useful
PT for preparing amphotericin derivative or analog antibiotic agent with
PT altered properties, in biosynthesis of polyketide other than
PT amphotericin.
XX
PS Claim 1; Page 52-114; 276pp; English.
XX

The invention relates to the gene cluster encoding the polypeptides
CC responsible for the biosynthesis of the polyene antibiotic amphotericin
CC (amph) of Streptomycetes nodosus. Polynucleotides of the invention are
CC useful for preparing amphotericin derivatives or analogue antibiotic
CC agents with altered properties and in the biosynthesis of polyketides
CC other than amphotericin. amphiIII, amphiII or amphiI mutants are useful
CC for producing amphotericin derivatives glycosylated with alternative
CC sugars; amphiDIII or amphiII gene sequences are useful in engineered
CC biosynthesis of perosaminyl-amphoteronolide B; amphiDII or amphiDI and
CC amphiN gene sequences are useful in the engineered biosynthesis of
CC perosaminyl-16-desacetoxy-16-methyl amphoteronolide B; amphiDIII, amphiDI
CC and amphiI gene sequences are useful for preparing polypeptides capable
CC of addition of mycosamine to a polyketide other than amphoteronolide A or
CC B or for preparing polypeptides for in vitro synthesis of GDP-mycosamine.
XX The present sequence is S. nodosus amphi biosynthetic gene cluster
```







This Page Blank (uspio)







```
Dh 259 GTCTGCGGGGGTATGCGGCCCGGTGCGGTGCTGCGGACACTTGTGCTCAAGAGC 318
Qy 307 GGCATGCGCGCGAGCTCGCGGCCACGCCCGCGGCTCCAGCAGTGGGCTCAACCG 366
Dh 319 CGGATGCGCGGGGGGCTGAGGGGGCGCCCGGGGTGAGGCGGTGGGGCTGAACCG 378
Qy 367 ATGTTGCCCCCGCGCGGATGACCGGCGAACCCGTGGCCCGCTGTGTCACCAAGGAC 426
Dh 379 ATGTTGCCCCCTGCTGAGGTCTTCAGGGGCGCGCGGTGCGGCGGTGTGTACCGAC 438
Qy 427 GGGCGGGGCGTCACGGCGCTGCTGCGGCTGTGAGGGCGCGCGCGCGAGCCGTACGG 486
Dh 439 GGGCGCGGTGTGCGGCGCTGTGAGCTGTGTGCGCGGAGTGGGAGCGCGGTGTGAG 498
Qy 487 CTCACGCGCGAGGAGCAGACCGGACGCGCGGACCCAGGCGCTTGAAGCAGCGCGTG 546
Dh 499 ATCCGCGCGCGGGGAGCAGACGAGCTGACCGCGCGGAGAGGCGCGACGACATGCGCG 558
Qy 547 CTCTCTCTTCTGCGGCTGCGCTGCGCGCGCTGCGGCGTGAAGTCCGAGCGCTGCGCGCG 606
Dh 559 GTCTGCGCTTTCGGGCTGCGGCTGAGCTGTGCGGTGAGCGTGGGGGCGCTGCGGGAC 618
Qy 607 AGCGACCGCGCGCGCCACGAGGTGCTGCGCGCTGCGCGCGCTGTGTGCTGCGCGGAC 666
Dh 619 AGTCCCGCGCGCGCTGTGAGCTGTGCGCTGCGCTGCGCTGCGATGCGCGGAGACG 678
Qy 667 CCCGAGGTATGAGGAGCATTCAGCGGTCCAAACCCCGCGCGCGCTGCGCGCGCGGCG 726
Dh 679 CGGAGGTATTTTTCAGACATTCAGGCGCGCAACCCCGCGCGCGCGCGCGGAGGCG 738
Qy 727 CTGCGGAGGCGCTGCGGCTCTTTCGCGCGCGCTGTGCGGAGCAGACCCGAGCGGTGCGAC 786
Dh 739 CTGCGCGCGCGCGCTGTGCGGCTGTGCGGAGCGGTGCGAGAGGAGCAGAGAGCTTC 798
Qy 787 GCGCCGCGCGCGCGCGACCGCCCGGCGCATCCGCGGAGAGCGACG 832
Dh 799 GCGGCGCTTTCGCGGACCTGCGGCGGTGTGCGGAGCAGCAGCGCG 844
```

## RESULT 2

```
US-09-987-614A-2
; Sequence 2, Application US/09987614A
; Patent No. 683382
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMBAS-JACOUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/09/987,614A
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 03-08/765,907
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 888
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
; US-09-987-614A-2
```

```
Query Match 27.9%; Score 270.8; DB 4; Length 888;
Best Local Similarity 58.8%; Pred. No. 9.3e-34;
Matches 486; Conservative 0; Mismatches 337; Indels 3; Gaps 1;
```

```
Qy 10 TTCCCGGAGCGCTGCTGTGCGCGGAGCGGGCGGTGGGCGCATGTT---CGCCGGG 66
Dh 19 TTCCGGGCGTTGTGTGTGTGTGGCGGGGCGGGTGGCGGTGGCGCATGTTCAAGCCACTTG 78
Qy 67 CTCTGCGGAGAGCGGGGAGCGGACAGCTGTGTGTGTGACTGTGTACCGCGCGGAGACGG 126
Dh 79 CTGTGTGCTTTCGGGGGTGGCGGTGTGACTGTGTGTGAGGTGGGCGGGGCGGTGTGGGAGAC 138
Qy 127 CCGGACGCTGTGCTGTGTGGAGCAGTCAACGCGCGGAGGCGCGAACTCGCGGCGCGCTTC 186
Dh 139 GGGGTGCGGGGTGTGTGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 198
Qy 187 CCGGACGCGGACCTGTCTGTCTGTGCGCGTACAGAGAGCTGTGCTTCAAGGCGGTGTGAG 246
Dh 199 GCGGCGCGGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 258
Qy 247 CCGGTGACCGCGGTCTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 306
Dh 259 GTGCTGTGCGGGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 318
Qy 307 GGCATGCGCGGAGAGCTGCGGCGCCACGCGCGCGCGGCTGTGAGAGCTGTGTGTGTGTGTGT 366
Dh 319 CGGATGCGCGGCGGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 378
Qy 367 ATGTTGCGCGCGCGCGCGCATGTGACCGGCGAGCCCGGTGCGCGCGGTGTGTGTGTGTGT 426
Dh 379 ATGTTGCGCGCGCGCGCGCATGTGAGGGGCGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGT 438
Qy 427 GGGCGGGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 486
Dh 439 GGGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 498
Qy 487 CTCACGCGGAGAGAGCAGACCGGAGCGGCGGACCCAGGCGCTTGAAGCAGCGCGTGAC 546
Dh 499 ATGCGCGCGCGCGCGGAGCAGACGCTGTGAGCGCGCGGAGAGGCGCGCATGCTGCGG 558
Qy 547 CTCTCTCTTCTGCGGCTGCGCTGCGCGCGCTGCGGCGTGTGAGCTGCGGCGCTTGTGCGGCG 606
Dh 559 GTGCTGTGCGCTTTCGGGCTGCGGCTGTGAGCTGTGCGGTGAGCGTGGGGGCGCTGCGGGAC 618
Qy 607 AGCGACCGCGCGCGCCACAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 666
Dh 619 AGTCCCGCGCGCGCGCATGTGCGATGTGCGGCTGTGCGGACCTGCGCGGAGACG 678
Qy 667 CCCGAGGTATGAGGAGCATTCAGCGGTCCAAACCCCGCGCGCGCTGCGCGCGCGGCG 726
Dh 679 CGGAGGTATTTTTCAGACATTCAGGCGCGCAACCCCGCGCGCGCGCGCGGAGGCG 738
Qy 727 CTGCGGAGGCGCTGTGCGCTCTTTCGCGCGCGCTGTGCGGAGCAGACCCGAGCGGTGCGAC 786
Dh 739 CTGCGCGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 798
Qy 787 GCGCCGCGCGCGCGCGAGCGCCCGGCGCATCCCGGGGAGTGGAGCG 832
Dh 799 GCGGCGCTTTCGCGGACCTGCGGCGGTGTGCGGAGCAGCAGCGCG 844
```

## RESULT 3

```
US-08-765-907A-1/c
; Sequence 1, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMBAS-JACOUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
```







QY	607	10GGACACGGCGCCCGACAGATGCTGTCCGCCCTCTGGCCGATGTCTCTGGCGGACAC	666
QY	1218	AGTGGCCCCGGCGCGCATCTGGCGATGTGGGGCTGTGGCGCCGGATGCGCGGGGACG	1159
QY	667	CCCGAGGTATACGGGGACATCATCAGCGGTCCAACCCCGGGCGGCGCTTCGCGCGCGCGGCG	726
Db	1158	CCGAGGTGTATTTTCGATCTCCAGGCGCGCCAAACCCCGGCGGCGCGCGCGCGGACG	1099
QY	727	CTCGCCGAGGCCCTTGCCTCTCTTCGCGCGCGCTGTGGCGAGCCGCGGACCTGTGCCGAC	786
Db	1098	CTGGGCGCGCGCTGTGTGCTGGCTGGCGGCGAGGCGCGTGTGAGAGGGGCGACGAGAGCGTTT	1039
QY	787	GCCCCCGGGCGCGGACGCCCCCGGCCATCCCGGGGATGCGAC	832
Db	1038	GCCGCGCTGTTCGCGGAACTGCGCGGTGTCTGGCGCGGACGCGG	993

## RESULT 5

```

1 Sequence 7. Application US/0804222C
2 Patent No. 5876991
3
4 GENERAL INFORMATION:
5 APPLICANT: DeHoff, Bradley S.
6 APPLICANT: Kuhstoss, Stuart A.
7 APPLICANT: Rostreck, Paul R., Jr.
8 APPLICANT: Sutcon, Kimberly L.
9 TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
10
11 NUMBER OF SEQUENCES: 15
12 CORRESPONDENCE ADDRESS:
13 ADDRESSEE: THOMAS G. PLANT 1501
14 STREET: LILLY CORPORATE CENTER
15 CITY: INDIANAPOLIS
16
17 STATE: IN
18 COUNTRY: USA
19
20 ZIP: 46285
21
22 COMPUTER READABLE FORM:
23 MEDIUM TYPE: Floppy disk
24 COMPUTER: IBM Compatible
25 OPERATING SYSTEM: MS-DOS
26 SOFTWARE: ASCII (DOS) Text only
27
28 CURRENT APPLICATION DATA:
29 APPLICATION NUMBER: US/08/804,227C
30 FILING DATE: February 21, 1997
31 CLASSIFICATION: 435
32
33 ACTORNEY/AGENT INFORMATION:
34 NAME: Plant, Thomas, G.
35 REGISTRATION NUMBER: 35,784
36 REFERENCE/DOCKET NUMBER: X-8231
37 TELECOMMUNICATION INFORMATION:
38 TELEPHONE: 317-276-2459
39 INFORMATION FOR SEQ ID NO: 7:
40
41 SEQUENCE CHARACTERISTICS:
42 LENGTH: 44377 base pairs
43 TYPE: nucleic acid
44 STRANDEDNESS: single
45 TOPOLOGY: linear
46 MOLECULE TYPE: DNA (genomic)
47
48 FEATURE:
49 NAME/KEY: CDS
50 LOCATION: 350..14002
51 FEATURE:
52 NAME/KEY: CDS
53 LOCATION: 14046..20036
54 FEATURE:
55 NAME/KEY: CDS
56 LOCATION: 20110..31284
57 FEATURE:
58 NAME/KEY: CDS
59 LOCATION: 31329..36071
60 FEATURE:
61 NAME/KEY: CDS
62 LOCATION: 36155..41830

```

US-08-804-227C-7

Query Match	10.8%	Score 104.4;	DB 2;	Length 44377;
Best Local Similarity	46.7%;	Pred. No. 2.9e-08;		
Matches 438; Conservative	0;	Mismatches 491;	Indels 9;	Gaps 3;

QY	22	GTGCTGCTGCGCGGCAAGCGGGGGGAGTGGGGGCGGATTTGTCGCGGGGCTGTGGCGGAGGGG	81
Db	29490	GTCTCTCGGTATATGCGATGAGCGGCGAGGGGGTGTCCATTGTCGGGGGCTGCGGTGGCCAG	294311
QY	82	GGCAGCGGCACGCGTGTGCTGTGACCTGTGATCCGCGCGCGGAGACGCGCGGACGCTGCTG	141
Db	29430	CACCTCCGGCC---CGAGGTGGGGGGGATTTGTGACACGGGGCATGTCGACGCCGCCCGGCG	293747
QY	142	GTGGGGCAGCTCACCGCGCTGGGGGCCGAACTTGGCGGCGCTCTCCGGGACGGGACCTC	201
Db	29373	GGCGGCTGTGACAGACGCCGCTGCCCGCGGCTTGAGCGCGGCGCAGGGTGAACGACCGTGA	293141
QY	202	GTCTGTGCTGCGCGGAC---CGAGGAGCTGGGCGCTTAAAGGCGGTGGCGCGCGGTGACCCGG	258
Db	29313	GTAGGCGCTAGAGAAACCGCGGGGAGCCGAGGGGGCTTGAGGATAGACACACCGTTCGG	292544
QY	259	CTCATGCGGCGGAGCGGCGTGTGCTGTGCGCGACACCTGTCTCCGTCCGAGCGGAGATGGC---	315
Db	29253	GACGGGGGGCGAGGAGCGCGGTGTGTGGGCGAAGCACAGCGGGGCCAAGGCCGCGCTCCACAG	291944
QY	316	GCGGAGCTCGCGGCGCCACGCGCCCGCGCTTCAGACAGTGGGCTTCAACCGATGTTGCGC	375
Db	29193	GCGAGGACCGCGGTGTCCCGACGCGGACACCGGTGAGCGCGGGGCGGACCTTGACAGACGGT	291344
QY	376	CCGCGCGCGGCGATGACCGGCGGACCCCGTGTGGCGCGCGGTGTGACACAGGAGCGGGCCGGCG	435
Db	29133	GCGGCGCGCTCTCGTTCCTTCCATTCTGCGGCTGTGGCGGGTACATGCGAGGGCGATGAGGGC	290744
QY	436	GTCACGCGCTTGTCTGCGGCTGTGTGAGAGGCGGCGGCGGCGAGGCGCGGTACGCTCACGGCG	495
Db	29073	GTGGCGGAATGTACCGCGCGGGCGGATACGGGAGCGGCACCTGGCGGGCTTCAGCGG	290144
QY	496	GAGAGGCAAGACCGGACGAGCGGCGGCGCACCCAGGCGCTGAGCGACGCGCTGCTTCTTCC	555
Db	29013	CGGCTCGGCGTCCGGGGGCGGCGACACAGGGCGAGGCGGTGAGGGTGGGTGCGTGCCTCCG	289544
QY	556	TTTGGGCTGTGCGCTTCCCGCGCGCTTCCGCGGTGACAGTCCCGGCGCTTGCGGCGACGCGACCG	615
Db	28953	GCCGAGCGCGCAGGACACCGTGTGTGGGCGAGGTGAAGCGGCGCTGTCCGTGTGCGCGCG	288944
QY	616	CCGCGCCACAGGTGTGCTGTGCGCTCTGTGCGCGCGGTGTCTGCGCGGACGCGCCAGGTTG	675
Db	28893	CCGCTTCAGCGGGGGGACGTAACCGTCCCGGCGCGGACCGGGGTCTGGGTTCGCCCGGA	288344
QY	676	TACGGGGAATCCAGCGGTTCACCCCGGGCGGCGTCTCGCGGCGCGGCGCTGCGGAG	735
Db	28833	GAGGAGGCGCTCGGTGATGAGCGCGCGGCGGACCGCGGCTGTGTGTCAGGTTCAAGAGAG	287744
QY	736	GGCCTGTGCTCTTCCCGCGCGGTGTGCGCGGACGACCGCGGACCGGTGCGGACGCGCCCGGG	795
Db	28773	GAAACCGGCGGGGCTCTTCTGCGCTGTGCGGAGGCGCAGACAGGCCCTCAACGCGGCGGCTCGGCG	287144
QY	796	CGCGCGACGCGCGCGCGGACATCCCGGGGAGATCGACGCGCGCGGAGAACTGTACGCGGCTC	855
Db	28713	CGGGTGTTCACCTTCGCGGTGTGTGCGGTATAGCGGACCGCGCGGGGTGGCCACACAG	286544
QY	856	TTTGGGGAATCTCGCGGCTCATTTGGACCGGAGCTTGTGCGCGCGGCGCAGGACCATGCGAG	915
Db	28653	GCGGGCGCGGTTCGGGCGCTTCGGGGGCGCGAGCGCGGCGGTGAGAGGTTCAGACCCG	285944
QY	916	GAGCTGTTCGACACCTTCACCGGACCGACGACGAGG	953
Db	28593	GCAAGCCACGATGTGTGCGGTGTGCGCGGCGACGTCGTGCGG	28556

## RESULT 6

US-08-804-198-1/c



```

Sequence 1 Application US/08804198
/ Patent No. 5945320
/ GENERAL INFORMATION:
/ APPLICANT: Burgett, Stanley G.
/ APPLICANT: Kuhseos, Stuart A.
/ APPLICANT: Rao, Nagaraja R.
/ APPLICANT: Richardson, Mark A.
/ APPLICANT: Roestek, Paul R., Jr.
/ TITLE OF INVENTION: PLATENOILIDE SYNTHASE GENE
/ NUMBER OF SEQUENCES: 6
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: PAUL R. CANTRELL 1138
/ STREET: LILLY CORPORATE CENTER
/ CITY: INDIANAPOLIS
/ STATE: IN
/ COUNTRY: USA
/ ZIP: 46285
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: Macintosh
/ OPERATING SYSTEM: Macintosh 7.0
/ SOFTWARE: Microsoft Word 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/804,198
/ FILING DATE:
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: CANTRELL, PAUL R.
/ REGISTRATION NUMBER: 36,470
/ REFERENCE/DOCKET NUMBER: P9113
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 317-276-3885
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 44377 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 350..14002
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 14046..20036
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 20110..31284
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 31329..36071
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 36155..41830
/ US-08-804-198-1

Query Match 10.8%; Score 104.4; DB 2; Length 44377;
Best Local Similarity 46.7%; Pred. No. 2,9e-08;
Matches 438; Conservative 0; Mismatches 491; Indels 9; Gaps 3;

OY 22 GTCGTCGTCGCGGCGACCGGCGCGTGGGCGGATGTCGCGGCGCTGCGCGGAGGCG 81
Db 29490 GTCGTCGATGATGCGCATGCGGCGGCGGCGGTCGCGGCGCTCGCGGTCGCG 294311
OY 82 GGCAGCCGCGACGCTGCTGCTGCTGACCTGTACCGCGCGCGGAGACGCGCGACGCTGCTG 141
Db 29430 CACCTCCGCGC---CGAGGTGGCGGCGGATCTGCACCGCGGCATGCGCGACGCCCCGCG 29374
OY 142 GTGGGCGACGTCACCGCGCGGCGCGGAGCTGCGGCGCGCTCCGCGGAGCGGACCTTC 201
Db 29373 GCGCGCGTGCACGAGCACCTCGCTCCCGCGCGCTGAGCGCGGACGAGGTGACGACCTGTA 29314
OY 202 GTCCTGCTCGCGCGTACA---CGAGAGCTGAGCCCTCAAGCGCTGAGCGCCCGCTGACCCG 258

```

```

Db      29313 GTAGGCGCTAGAGAAACAACCGCGGGGAGACCGAGCGGCGCTGGGGGTAGAGACACACCGCTCCGG 29255
Oy      259 CTCATGCGAGCGGGGCGCGCTGCTCGCCGACACCTGTCTCCGTCCGACGGGGCATGGC---C 315
Db      29253 GACCGGGGCGAGAGACCGGAGTGTGGCCACACGACAGCGGGGCCACGGCCCGCTCCACAG 29194
Oy      316 GCGGAGTTCGGGGCCCAACGCCCCCGGCGTTCACAGACATGTGGGCTCAACCCGATGTTTCGC 375
Db      29193 GCGGAGACACCGGTGTCCCGACCGGCGACACCGGTGAGCGCGGGGCGCATCTGACAGCGGT 29134
Oy      376 CCGCGCGCGCGGCATGACCGGCGGACCCGCTGAGCGCGCGCGCTGTGTACACGAGGACGGGCGGGC 435
Db      29133 GCGGCGCGCTCCGTTCCATCTCCGCTCTCGCCCGGATACATGCCAGGGCCATAGAGGC 29074
Oy      436 GTCAACGCGCCCTGTGCGGCTGTGTGAGGGCGAGGAGGCGGAGGCGCGTACGAGTCAAGCGC 495
Db      29073 GTGCGGGAAGTTGACGCGCGCGCGCGGTACGGGACGCGCATCTGGCCGGGCTCCAGCGG 29014
Oy      496 GAGAGACACGACCGGACGAGCGGCGACCCAGGCGCTTACGACGCGCTGCTCTCTTC 555
Db      29013 CGCTTCGGCTTCGCGGGCGGGGACCGAGGCGAGGCGGTGAGAGGTAGGTGCGTGCCTCG 28954
Oy      556 TTGCGGCGTGCCTTCGCGCGCGCTCGGCGGTGAGCGTCCGGGCGCTGAGCGGCGACCGCACCG 615
Db      28953 GCCAGCGCGCACGACCGTGTGCGGGCGGGGTGAGCGGCGCGTCCGTGCGGTGCGCGGCG 28894
Oy      616 CCGCGCCACACGAGTGTGCTCGCTCTGAGCCCGTGTGCTCGCGCGGACGCGCGAGGTG 675
Db      28893 CCGCTCCAGCGGGGCGACGTACACCGTCCCGGCGCGGACCGGAGTGTGCGGTTCGCGGGA 28834
Oy      676 TACGGGACATCCAGCGGTCAACCCCGGCGGGCGTCCGCGCGCGGGCGCTCGCCGAG 735
Db      28833 GCGAGGGCGTCCGTGACGCGCCCGGCGGAGCGCGGAGTGTGATCAGGTTCACGAGCAG 28774
Oy      736 GCCCTGCGCTCTTCGCGCGCGGTGTGCGGAGACCGGGAACGTACCGAGCGCCCGCGG 795
Db      28773 GAACCGGCGCGGCTCTTCGCGCTGCGCGGAGCGACGACGACCGCCAGCGCGCGGCG 28714
Oy      796 CGCGCCACGCGCCCGGACATCCCGGGGATGCGAGCGCGCGGACCTTCAGCGGCTG 855
Db      28713 CGGTCGTCCACTTCGCGGTGTGCGGTACGAGACCGCGCGCGGAGTGGCCACACAG 28654
Oy      856 TTCGGGGAACCTCGCGGCTCATGGACCGGAGCTGCGCGCGGGCCAGACCACTGCGAG 915
Db      28653 GCGGCGCGCTCGGGCGCTCCGGGCGCGGAGCGCGGCTGTGAGAGGTCCAGCACCG 28594
Oy      916 GAGCTGTTCCGACCTTCACCGGACCGGACCGACGAGGAG 953
Db      28593 GCAGGCCACGAGTGTGCGGTGCGCGGCGGACGTGCGCG 28556

RESULT 7
US-09-105-537-30
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-105-537-30

```



```
Query Match      10.7%; Score 104; DB 3; Length 13842;
Best Local Similarity 45.1%; Pred. No. 3,7e-08;
Matches 425; Conservative 0; Mismatches 515; Indels 2; Gaps 1;

QY 27 CCGCGGCGGAGCGGGGCGGTGGCGGCGCATGTTCCGCGGCGTCTGTGGGAGCGGGCGAG 86
DB 12219 CGACGGACCATCATGAGGTGGCGCGCGCTGGGCTCCGTCCGCGCGGAGCCGCGACCGCGCG 12278
QY 87 CCGCACGCTGCTGTCGACCTCTGTAACCGCGCGCGGAGCGGACCGCTGCTGTGGG 146
DB 12279 ACCGGGCGTGGCGCGCGGAGGAGGACGAGTCTGTGACCGCGGCGGACCGGCGGCGG 12338
QY 147 CGAGCTCACCGCGCGCGGCGCGGCGGAACTCGCGCGCGCGCTTCGCGGAGCGCGACCTGCTCT 206
DB 12339 ACTGATCGCGCGGCGACGCTGTGGGCGAGTGGGCGGTACAGCGCTGCTGTGTGAGCGCG 12398
QY 207 GCTTCGCGGTACAGAGGAGACGTGGCCCTCAAGGCGGTGGCGCGCGCTGACCGCGCTCATGCG 266
DB 12399 GCGGGGACGAGAGCGCGCGCGCGCGCGCGCGAGCTGTGACAGAGCTGAGGCGCTGGGAGCG 12458
QY 267 GCGGCGCGCGCTGCTGCGCGACACCTGTCTCGGACGCGGCGATGGCGCGGAGCTGCG 326
DB 12459 CGAGCTTCGCTGGCGCGCGCGGTGCGACGTGCGCGAGACCGGAGCGCTTACCGCGCTACTCGA 12518
QY 327 GCGCCACGCGCGCGCGCGGTCCAGCACTGGGCGCTTCAACCGATGTTGCGCGCGCGCGCG 386
DB 12519 CGCGCATCCCGCGGAAACCGCGCTCACCGCGGTGTCTACACAGCGAGGCGCTCTCTTCGA 12578
QY 387 CATGACCGGCGCGACCGCGTGGCGCGCGGTGTACACAGGAGAGGCGCGCGCGCTTACCGCGCT 446
DB 12579 CGGACCTCTCCGCTCCATGACAGAGGAGAGCTGGAACACGTACTGTGCGCGCGCAAGGTGCA 12638
QY 447 GCTGCGGCTCTGTGAGGAGCGGCGCGCGCGCGCGCTGACGCGCTACAGCGAGAGACGCA 506
DB 12639 CGCGCGCTTCTCTCTGACAGAACTCACTGTGACGCGCGCGCTAGACCTGTGGACGCTTCTGT 12698
QY 507 CCGGACGAGCGGCGCGCAACCGAGCGCTTACGACGCGCGTCTCTCTTGTGGGCTGCG 566
DB 12699 CATGTTCTCTCTCGCGCGCGCGCGCTTCTGTGGGCGCGGCGAGGCGCGCTTACCGCGCGCG 12758
QY 567 CTTGCGCGCGCTGTGGGCTGACAGTTCGGGCGCGCTGTGGCGGCGACCGGCGCGCGCGCGCA 626
DB 12759 CAAGCGCACCTTGCAGCGCGCTGTGGCGCGCGCGCGCGCGCGAGCTTCCGCGCGCTTCTC 12818
QY 627 GGTGCTGCTGCGCGCTCTGTGGCGCGGTGTGCTGTGGCGGAGCGCGCGGTGTACGAGGACAT 686
DB 12819 CTTGCGGTGGGCGCTTGTGGGCGCGAGCGACGCGCATGACCGGCGAGCTTGGCGACGCGGA 12878
QY 687 CCAAGCGGTCCAACCGCGCGCGCGCGGTGCGCGCGCGCGCGCGCTGCGCGAGCGCTTGCCTTC 746
DB 12879 CTTGCGCGCGGTATGAGCGCGCGCGCGCATGCGCGGATACAGGACGCGAGGCGCATTCGCGCT 12938
QY 747 CTTTCCGCGCGCTGTGGCGGACGACCGCGGACCGGTGCGGAGCGCGCGCGCGCGCGCGCGCA 806
DB 12939 CTTGAGCGCGCGCGCTTGTGGCGGACGCGCGCGCGGTGTGCTGCTGCGCGCTTGTGCGCTGCA 12998
QY 807 CCGCGCGCATTCGCG--GGGATGCGAGCGCGCGCGGAACTTCAGACGCGGTCTTTCGGGAA 864
DB 12999 CGCGCGGCTGGGAGCGCGCGCGGAGACGACCGGCGGAAATCCGCGCGCTTTCGCGGA 13058
QY 865 CTCGCGCGGCTCATGGGACGAGGCTTGGCGCGGCGCGGAGCACTTTCAGAGAGCTTTC 924
DB 13059 CGTGTGTGGGCGGACGCTTCCGCGCGCGCGCGCGCTTTCGCGCGCTTTCGAGACAGC 13118
QY 925 CGGACCTTTCAGCGGACCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 966
DB 13119 CGGAGCGCGCGGACGCGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 13160
```

RESULT 8  
US-09-105-537-5  
; Sequence 5, Application US/09105537A  
; Patent No. 6265202

```
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-5

Query Match      10.7%; Score 104; DB 3; Length 36778;
Best Local Similarity 45.1%; Pred. No. 3,4e-08;
Matches 425; Conservative 0; Mismatches 515; Indels 2; Gaps 1;

QY 27 CCGCGGCGGAGCGGGGCGGTGGCGGCGCATGTTCCGCGGCGTCTGTGGGAGCGGGCGAG 86
DB 13960 CGACGGACCATCATGAGGTGGCGCGCGCTGGGCTCCGTCCGCGCGGAGCCGCGACCGCGCG 14019
QY 87 CCGCACGCTGCTGTCGACCTCTGTAACCGCGCGCGGAGCGGACCGCTGCTGTGGG 146
DB 14020 ACCGGGCGTGGCGCGCGGAGGAGGACGAGTCTGTGACCGCGGAGACCGGCGGCGGCGG 14079
QY 147 CGAGCTCACCGCGCGCGGCGCGGCGGAACTCGCGCGCGCGCTTCGCGGAGCGGAGCTGCTCT 206
DB 14080 ACTGATCGCGCGCGCACGTTGTGGGCGAGTGGGCGGTACAGCGCTGCTGTGAGCGCG 14139
QY 207 GCTTCGCGGTACAGAGGAGAGTGGCCCTCAAGGCGGTGGCGCGCGCTTACCGGCTCATTGG 266
DB 14140 GCGGGGACAGAGCGCGCGCGCGCGCGAGCTGTGACAGAGCTGAGAGCGCTTGGAGCG 14199
QY 267 GCGGCGCGCGCTGCTGTCGCGGACACCTGTCTCGTTCGCGAGCGGCGATGGCGCGAGCTGCG 326
DB 14200 CGAGCTTCGCTGGTGGCGCGCGGTGCGAGCGTGCACGCGGAGGCGCTTACCGCGCTACTCGA 14259
QY 327 GCGCCACGCGCGCGCGCGGTCCAGCACTGGGCGCTTCAACCGATGTTGCGCGCGCGCGCG 386
DB 14260 CGCGCATCCCGCGGAAACCGCGCTTACCGCGGTGTCTACAGGAGGCGGTCTCTTCGGA 14319
QY 387 CATGACCGGCGGACCGCGTGGCGCGGTGTGACAGGAGCGGCGCGGCGGTCTTACCGCGCGCT 446
DB 14320 CGGACCTTCTCGTTCATGACAGAGGAGAGCTGGAACAGTATGCGGCGCGCAAGGTGCA 14379
QY 447 GCTGCGGCTCTGTGAGGAGCGGCGCGGAGCGCGCTTACAGGCTTACAGCGGAGAGGAGGAG 506
DB 14380 CGCGCGGCTTCTCTGAGAGAACTCACTGAGAGCGCGCGGATGAGACTTGGCGAGCTTCTGT 14439
QY 507 CCGGACGAGCGGCGGCGCAACCGAGCGCTTACGAGCGCGGTGTCTCTTGTGGGCTGCG 566
DB 14440 CATGTTCTCTCTCGCGCGCGCGGTCTTGTGGGCGGCGGAGGCGCGCTTACGCGCGCG 14499
QY 567 CTTGCGCGCGCTGTGGGCTGAGAGTTCGGGCGCGTGTGGGCGGAGGAGGAGGAGGAGGAGGAGGAG 626
DB 14500 CAAGCGCACCTTGCAGCGCGCTTGTGGGCGCGCGCGGAGCGGAGCTTCCGCGCGCTTTC 14559
QY 627 GGTGCTGCTGCGCGCTTGTGGGCGCGGTGTGCTGTGGGCGGAGCGCGCGGTGTACGAGGACAT 686
DB 14560 CTTGCGGCTGGGCGCTTGTGGGCGGAGACGAGCGGCGGATGAGCGGAGCTTGGCGGAGGAG 14619
QY 687 CCAAGCGGTCCAACCGCGCGCGCGGTGCGCGCGCGCGCGCTTTCGCGGAGCGCTTGCCTTC 746
DB 14620 CTTGCGCGCGGTATGAGCGCGCGCGGAGATCAGCGAGGAGGAGGAGGAGGAGGAGGAGGAG 14679
QY 747 CTTTCCGCGCGCTGTGGGCGGAGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 806
DB 14680 CTTGAGCGCGCGCTTGTGGGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 14739
```



QY	807	CCCCGGCCATCCG--GGGATGGGACGGGCGCCGGGAACTCGACGGCGCTTCGGGGAA	864
Db	14740	CGCCGGCGTCGGGACGGCGCGCGGGAACGACCCGGCCGGAATCCCGCGCTTCCGGGA	14799
QY	865	CTCCGCCGCGCTCATGGAGCCGGAGCTGTCGCGCGGGCCAGGACCACTGCGAGAGCTGTC	924
Db	14800	CGTGTGTGGGCGCCAGGACCGTCCGGGCGCCGGCGCTTCGCGGCTTCGACGACAGC	14855
QY	925	CGCACCTCTCAACCGCACCGACGACGAAAGCGCAGAGAGACCGA	966
Db	14860	CGGGACGGCGCGGACGCGCCGGGAGCGGCGGACGCGCGGCGGA	14901

## RESULT 9

```

US-09-320-878-19
: Sequence 19, Application US/09320878A
: Patent No. 6117659
: GENERAL INFORMATION:
: APPLICANT: ASHLEY, Gary
: APPLICANT: BETLACH, Melanie C.
: APPLICANT: BETLACH, Mary C.
: APPLICANT: MCDANIEL, Robert
: APPLICANT: TANG, Li
: TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
: FILE REFERENCE: 30062200210
: CURRENT APPLICATION NUMBER: US/09/320, 878A
: CURRENT FILING DATE: 1999-05-27
: EARLIER APPLICATION NUMBER: CIP OF 09/141, 908
: EARLIER FILING DATE: 1998-08-28
: EARLIER APPLICATION NUMBER: CIP OF 09/073, 538
: EARLIER FILING DATE: 1998-05-06
: EARLIER APPLICATION NUMBER: CIP OF 08/846,247
: EARLIER FILING DATE: 1997-04-30
: EARLIER APPLICATION NUMBER: 60/119, 139
: EARLIER FILING DATE: 1999-02-08
: EARLIER APPLICATION NUMBER: 60/100, 880
: EARLIER FILING DATE: 1998-09-22
: EARLIER APPLICATION NUMBER: 60/087, 080
: EARLIER FILING DATE: 1998-05-28
: NUMBER OF SEQ ID NOS: 34
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 19
: LENGTH: 38506
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-320-878-19

```

	Query Match	Similarity	10.7%	Score	104	DB	3	Length	38506
	Best Local	Similarity	45.1%	Pred.	0.3	38-08			
	Matches	425	Conservative	0	Mismatches	515	Indels	2	Gaps
Qy	27	CGTCGCGCGACGCGGCGCGTGGCGGCGCATGTCGCGGCTGTCGCGAGGCGGCGACG	86						
Db	12102	CGACGGGACCATCAGGCTGGCGCCGCTGGGCTCGGCTCGTCCGCGACCGGACCGCCCGC	12161						
Qy	87	CCGCAAGCTGTCGTCACCTGTCACCGCGCGCGGACCGGCGCGGACCGCTGCGTGGG	146						
Db	12162	ACCGGCGTTCGCCCGGAGGGCACAGTTCGTGACCGGCGGACCGCGCGGCTTGGGCGG	12221						
Qy	147	CGACGTACCGCGCGCGGCGCCGACATCGCGGCGCGCTTCGCGGACCGGACCTCTGCT	206						
Db	12222	ACTGTCGCCCCGCGACGTGGTGGGCGAGTGGGGCTGACGCTGCTGCTGGTAGCGG	12281						
Qy	207	GCTGCGCGTACCGAGGACGTGGGCTCTCAAGGCGCTGAGGCGCCGTCACCGGCTCATGG	266						
Db	12282	GCGGAGCACCGACGCCCGCGGCGCGACGACTGTGTCACGACCTGAGGCTCTGGAGAC	12341						
Qy	267	GCCGAGGCGCGCTGCTGCGCGACACCTGTCGCTCCGACCGGCGATGGCGCGAGCTCGC	326						
Db	12342	CGAGGTCTCGGTGGCGCGGTGCGACGTCGCGACCGGAAAGCCCTTCACGCGCTACTCGA	12401						
Qy	327	GGCCACGCGCCCGCGCGTCCAGCACGTTGGGCTTCATCCGATGTTGCGCCCGCCGCG	386						

Db	12402	CGCATTCCCGCGGAACACCCGCTCACCGCGGTGTCACAGCGGACGAGCGTCTCTCCGA	12461
OY	387	CATGACCGGGCCAAACCCGTGGCCGCGCTGTCAACAGGAGCGGGCGGTCAACGGCCCT	446
Db	12462	CGGACACCTCCCGTTCATGACGACGGAGGACGTGGAAACATGATCGGCCCCCAAGTTCGA	12521
OY	447	GCTGCGGCTGTGTGAGAGGCGCGCGGCGGACAGCCCTGTAACGGCTCAACGGCGGAGACGGA	506
Db	12522	CGCCGCGTTCCTCTCTGACGGAATCAACTCTGAGGCCCGGATADGACTGTGGACGTTGCT	12581
OY	507	CCGAGACGACGCGGAGCCACCCAGGCGCTTGACGACGCGGTCTCTCTTGGGCTCGC	566
Db	12582	CATGTTCTCTCCGCGCGCGCGCTGTCTGAGTGGCGCGGGACAGGCGGCTTACGCGCGCG	12641
OY	567	CTTCGCGCGGCTGTGGGTGAGACGTCCGCGGCGCTGTGGGGGAGACGGACCGCGGCCACCA	626
Db	12642	CAAGCGCACCTTGACGCGCCTCGCTGTGGGCGCGCGCGGAGCGGAGCTCCCGCCTCTTC	12701
OY	627	GGATCTGCTGACCTCCCTCGTGGCCGCTGTGCTGTGGCGGACGCCCGAGAGTTACGGGGACAT	686
Db	12702	CTTCGCGTGGGGGCTCTGTGGGCGGACACAGCGGCATGACCGGGAGCTGGCGACGGCGGA	12761
OY	687	CGAGCGGTCCAAACCCCCCGGCGGCGGTGCGCGCGCGCGGCGCTGTGGCGAGGCCCTTGCGCTC	746
Db	12762	CTGTGCGCGGATGACGCGCGCGGGACATCGGCGGGAATCAACGACGCGCGAGGGCATGCGCT	12821
OY	747	CTTTCGCGCGCGCTGTGGGCGGACGACCCCGGACCGTGTCCGACGCCGCCCGGCGCGCGGACGC	806
Db	12822	CTTGACGCGCGCCTTCGCGACACCGCACGCCGCGGTCTGTGTGCCCTTGCGGCTCGACGC	12881
OY	807	CCCCGGCCATGCCG--GAGGATGCGACGCGCGCGGAGAACCTTGACGCGCTCTTTCGGGGAA	864
Db	12882	GGCGGGGCTGCGGGAAGCGCGCGGGGAACGACCCGCGGGAATCCGGGGCTCTTCCGGGA	12941
OY	865	CTTCGCGGCTCATGTGGGACCGGAGCTTGTGGGCGGGGCGGACGACCTGTGCCAGAGCTGTTTC	924
Db	12942	CGTGTGTGGGCGCCAGGACCGTCTCGGGCGCGGCGGTCTCGCGGCTCTCGCTTACGACMAC	13001
OY	925	CGACCCCTTCACCGGACCGACGGAAGGCGGAAAGACGGA	966
Db	13002	CGGAGCGGCTCGGACGCGCGGGGACGGCGGAGCGCGCGCGGA	13043

## RESULT 10

US-09-141-908-1  
Sequence 1, Application US/09141908  
Patent No. 6503741  
GENERAL INFORMATION:  
APPLICANT: ASHLEY, Gary  
APPLICANT: BETLACH, Melanie C.  
APPLICANT: BETLACH, Mary  
APPLICANT: MCDANIEL, Robert  
APPLICANT: TANG, Li  
TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a  
TITLE OF INVENTION: Modular PKS Gene Cluster as Scaffold  
FILE REFERENCE: 300622002100  
CURRENT APPLICATION NUMBER: US/09/141,908  
CURRENT FILING DATE: 1998-08-28  
EARLIER APPLICATION NUMBER: CIP OF 09/073,538  
EARLIER FILING DATE: 1998-05-06  
EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
EARLIER FILING DATE: 1997-04-30  
EARLIER APPLICATION NUMBER: PROV. 60/076,919  
EARLIER FILING DATE: 1998-03-05  
EARLIER APPLICATION NUMBER: PROV. 60/087,080  
EARLIER FILING DATE: 1998-05-28  
NUMBER OF SEQ ID NOS: 31  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 1  
LENGTH: 38506  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-141-908-1



Query Match	10.7%	Score 104;	DB 4;	Length 38506;
Best Local Similarity	45.1%;	Pred. No. 3.3e-08;		
Matches 425;	Conservative 0;	Mismatches 515;	Indels 2;	Gaps 1;

OY	27	CGTGGCGGAGACCGGGGCGGTGGGCGGACATGTTGACCGGCTGCTGCGGAGAGCGGGCAG	86
Db	12102	CGAGGGACCAATCAAGGTGGGCGCCGCGTGGGCTCCGTCCGCGCCGAGACCGGACCCGCGC	12161
OY	87	CCGACGCTGTCGTGACCTCTGACCTGTAACCGCGCGCGGGACCGGCGGACCGCTGCTGGGG	146
Db	12162	ACCGGCGCTGCGCCCGGAGGGGACAGGTCTGTGTGACCGGCGGACCGGCGGCTTGGGCGG	12221
OY	147	CGAGTGAACCGCGCGGGGCGCCGAATCGCGGCGCCCTTCGGGACGCGGACCTCTGTCT	206
Db	12222	ACTGTCGCGCGGACGTGGTGGGCGAGTGAGGGGTGTGACGCGCTGCTGTGGTGAACCG	12281
OY	207	GCTGCGGCTACGAGAGACGTGGTCCCTTCAGGCGCGTGTGGCCCGGTGACCGGCTCATGGC	266
Db	12282	GCGGGGACGGAACCGCCCGGGCGCCGACGAGCTGTGTCAAGACTGAAGGCTCTGGAGAC	12341
OY	267	GCCGGGCGCGCTGTGCGCGACACCCGTGTCGTCGGAAGGGGATGAGCGCGGAGGTTCG	326
Db	12342	CGAGCTCTCGGTGGCGCGGTGTGCGACGTCGCGACCGGAAAGCTTCAACGCGGTACTTCA	12401
OY	327	GGCCACGCGCCCGCGGCTTCAGACGAGCTGGGCTTCAACCCATGTTGCGCCCGCGCGG	386
Db	12402	CGCATCCCGGCGGACACCGCGCTCACCGGCGGTGTGTCAACGCGAGGGTCTCTTCGA	12461
OY	387	CATACCGGCGGACCCGTGGTGGCGCGGTGTGTACAGAGGATGGGCGCGGCGTCACGGCT	446
Db	12462	CGGACCTCTCCGTCTCATGACGAGGAGGACGTGTGAACATGTCTCGGCGCCGAAGTCTCA	12521
OY	447	GCTCGGCTGTCGAGGAGCGGCGGCGGACGCGCTTACGCGCTACAGCGGAGGAGACGA	506
Db	12522	GCGCGGCTTCTCTTCAAGAACTCACTGACGCGCCGCAATACACTGTGACGACTTTCGT	12581
OY	507	CGGACGAGCGGCGGACCCAGGCGCTTGAAGCAAGCGGTTCCTCTTGGGCTCGC	566
Db	12582	CATGTTCTCCGCGCGCGCGCGCTTTCGATGGCGGCGGAGGAGCGCTTACGCGCGCG	12641
OY	567	CCTGCGCGCTCGGCGGTGACGTGTGGGCGCTGTGGGCGGACCGGACCGCGCTTACCA	626
Db	12642	CAAGGCAACCTTCAGACCCCTTCGTGTGGCGCGCGCGGCGAGCGGACTCCCGCTTTC	12701
OY	627	GGTCTGTGCGCCCTTCGAGCCGCTGTGTGTGCGGCGAGCGCCGAGATGTACGGGACAT	686
Db	12702	CCTGGGCTGGGCGCTTCTGGGCGCGAGACGAGCGGATGTACTGGGAGCTCGGCTAGCGGA	12761
OY	687	CGAGCGGTCAACCCCGGCGGCGGTTCGCGCGCGCGGCGGCTTCGCGAGGCGCTTGGCTTC	746
Db	12762	CTGCGCGGATGTAGCGCGCGGCGCATTCGGCGGGATATAGGAGCGCGGAGGATCGCGT	12821
OY	747	CTTGGCGCGCTGTGTGGCGGACGACCCCGGACCGTGTCCGACGCGCCCGGCGCGCGACGC	806
Db	12822	CCTTGAAGCGCGCTTCCTCGGACGACCCGCAACCGGATCTGTCTCCCTGTGGGCTTCAAC	12881
OY	807	CGCGGCGCATCCG--GGGATGTGAGGCGGCGCGGGAACCTGACGCGCGCTTTCGGGGAA	864
Db	12882	GCGCGGCGTGGGACGCGCGCGGGAACGACCCGCGCGAATCCGCGCTTCTTCGGGGA	12941
OY	865	CTTCGCGGCTCATGGAGCGGAGCTTCGCGGCGGCGCGGACCACTTCAGAGAGCTGTTC	924
Db	12942	CGTGTGCGGCGGAGACCGTTCGGGCGCGCGGCGCTTCGCGCTTCGCGCTCAAGCAAC	13001
OY	925	CGACACCTTCCACCGGACCGACGAGGAGGCGGAGAGACCGA	966
Db	13002	CGGAGCGGCGGACGCGCGGAGACCGGAGGACCGGCGGCGGA	13043

```

: Patent No. 6509455
: GENERAL INFORMATION:
: APPLICANT: ASHLEY, Gary
: APPLICANT: BETLACH, Melanie C.
: APPLICANT: BETLACH, Mary C.
: APPLICANT: MCDANIEL, Robert
: APPLICANT: TANG, Li
: TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
: FILE REFERENCE: 300622002120
: CURRENT APPLICATION NUMBER: US/09/657,440
: CURRENT FILING DATE: 2000-09-07
: PRIOR APPLICATION NUMBER: 09/320,878
: PRIOR FILING DATE: 1999-05-27
: PRIOR APPLICATION NUMBER: CIP OF 09/141,908
: PRIOR FILING DATE: 1998-08-28
: NUMBER OF SEQ ID NOS: 34
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 19
: LENGTH: 38506
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
: US-09-657-440-19

```

Query Match	10.7%	Score 104;	DB 4;	Length 38506;
Best Local Similarity	45.1%;	Pred. No. 3.3e-08;		
Matches 425; Conservative	0;	Mismatches 515;	Indels 2;	Gaps 1;

RESULT 11  
US-09-657-440-19  
; Sequence 19, Application US/09657440

27 CATTGCGGGAAGCGGGCGGTGGGCGGCATGTTTCCCGGGCTGTCTGGGAGGCGGCAG 86  
 12102 CGAGGGACCATCAGGTGTGGCCCGCTGGCGCTCCGTCGACCCGAAACCGGACGCGCCG 12161  
 Oy CCGCAGCTGTGTGTGTCGACCTCGTACCGGCGCCGGAGCGGCGCGCTGGCTGGGG 146  
 Db 12162 ACCGGCGCTGCCCGGAGGGACAGGATCTGTCTGACCGGGCGGCACCGGGGCGCTGGGCGG 12221  
 Oy 147 CGAGCTCACCGCGCCGGGGCCGGAATCGCGGCCTTCGCGGACGCGGACCTGTCT 206  
 Db 12222 ACTGTGCGCCCGGACAGGTGTGGGCGAGTGGGGCGTACGACGCTGCTGTGTAGCCG 12281  
 Oy 207 GCTTGGCGGTACAGAGGAGTGTGGCCCTTCAAGGCGGTGGGGCCCGTACCGGCTCATGGC 266  
 Db 12282 GCGGGGACGAGAGCGCCCGGGCGCCACAGAGCTGTGTGACAGAGCTGGAGGCTTGGGAGC 12341  
 Oy 267 GCCGGGCGCGCTGTGTGCGGCACACCTGTGTCCGTCCGGAGGGGACATGGCGCGAGCTGCG 326  
 Db 12342 CGAGGTCTGGTGTGGCGCGGTGGCAGCTGTCCGCCACCGCGAAGCTTCACCGCGCTACTCGA 12401  
 Oy 327 GGGCCACGCCCCGGCGGTCCAGCAGTGGGCTTCAACCCGATGTTTCGCCCGCGCGCG 386  
 Db 12402 CGCATTCGCGCGGAACACCGGCTCACCGGGGTGTCCACACGCGAGGGGTCTCTCGGA 12461  
 Oy 387 CATGACCGGGCGAACCCGTGGCGCGCGCTGTGTACACAGAGGACGGGCGGGGTCAAGGCTT 446  
 Db 12462 CGGATCCCTCCGTGTCAATGACGAGGAGGAGCTGGAAACATGTGCGGCCCAAGGTTCGA 12521  
 Oy 447 GCTCGGCTGTGTGAGGGGCGGCGGCGGAGCCGTAAGGCTTCAAGCGAGAGAGACGA 506  
 Db 12522 CGCGCGTTCCTCTCTGACGAACTACCTCGAGCCCGGATACGACTGTGGACGTTTGGT 12581  
 Oy 507 CCGGACGACGAGCGGACCCAGGCGCTTGACGACGCGCTGTCTTCTCTTGGGCTGCG 566  
 Db 12582 CATGTTCTCTCCGCGCGCGCGCTTCTTGAGTGGCGGGGAGAGGCGCTTACGCGCGCG 12641  
 Oy 567 CTTGCGCCCGCTGGGTGTGACGTCCGGGCTCTGGGGGGAGAGGACCGCGCCCAACA 626  
 Db 12642 CAAGCGCACCTTGACGCGCTTCCTGTGGGCGCGCGCGGGAGCGGACTTCCCGCTTCTC 12701  
 Oy 627 GGTGCTCTGCGCTCTGAGCCGCTGTGTGCTCGAGCGGACGCCCGAGGTGTACGGGACAT 686  
 Db 12702 CTTGCGGTGGGAGCTCTGGGCGAGAACAGCGGACATGACCGGCGAGCTGTGCGCAGGCGGA 12761  
 Oy 687 CCAAGCGTCCAAACCCCGGCGGCGTGTGCGCGCGCGGCGGCTGTGCGCGAGGCCCTTGCGTTC 746



Db 12762 CTGCGCCGAGATAGCCGCGGAGCATCGGCGGATCAGGACCGGAGCATCGGCT 12821  
QY 747 CTTCGCGCGCTGTGCGGAGACCGGACCGTGCAGCCGCCCGGCGCGCGGAGCG 806  
Db 12822 CTTGAGCGCGCCCTCGCGACGACCGCCACCGGCTCTCTGCTGCGGCTCGAGCG 12881  
QY 807 CCCCAGCATCCG--GGGATGCGAGCGCGCGGAGAACCTCGACGCGCTTTCGAGGAA 864  
Db 12882 CGCGGGGCTGCGGAGAGCGCGCGGAGAACGACCGCGCGGAAATCCGCGCTCTTCGAG 12941  
QY 865 CTCGCGCGGCTCATGGAGCGGAGCTGCGCGGCGGAGGACCACTGCGAGCTGTC 924  
Db 12942 CGTGTGCGGCGGAGGACCTCGCGGCGCGCGGCGCTCGCGCTCGCGCTCGAGAGCG 13001  
QY 925 CGGACCTCGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 966  
Db 13002 CGGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 13043

## RESULT 12

US-10-237-551-193  
; Sequence 193, Application US/10237551  
; Patent No. 6821519  
; GENERAL INFORMATION:  
; APPLICANT: Day, Craig H.  
; APPLICANT: Hosken, Nancy A.  
; APPLICANT: Parsons, Joseph M.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND  
; FILE REFERENCE: 210121.538C3  
; CURRENT FILING DATE: 2002-09-06  
; NUMBER OF SEQ ID NOS: 254  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 193  
; LENGTH: 3957  
; TYPE: DNA  
; ORGANISM: HSV2  
US-10-237-551-193

Query Match 10.6%; Score 102.6; DB 4; Length 3957;  
Best Local Similarity 46.8%; Pred. No. 6.9e-08;  
Matches 439; Conservative 0; Mismatches 484; Indels 16; Gaps 3;

QY 7 GGGTTCCTCCCGGAGCGTCTGCTGCGGCGGAGCGGCGGCTGCGGCGGATGTCGCGCG 66  
Db 1695 GCG 1752  
QY 67 CTGCTCGGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 126  
Db 1753 GCGGTTCG 1812  
QY 127 CCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 186  
Db 1813 GCGCGCGCGCTGCG 1872  
QY 187 CCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 246  
Db 1873 GGTGCGGCG 1932  
QY 247 CCGGTGACCGGCTCATGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 306  
Db 1933 CCGGAGATCTGAGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1992  
QY 307 GCGATGCGCGGAGCTGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 366  
Db 1993 CTGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 2052  
QY 367 ATGTTGCG 426  
Db 2053 CCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 2112  
QY 427 GGGCGCGGCGTCAAGCGCGCTGCTGCGGCGTGTGAGAGGCGGCGGCGGAGCGCGGAG 486

Db 2113 GCGCTGTGCTGATGCGCTGCGCGGAGGAGCTGCGGCTGCGCGGCGGAGCGGCGCG 2172  
QY 487 CTCACGCGGAG 546  
Db 2173 GTGCGCGCGTGTGCGCGCGTGTGAGCGCTGTGCTGTGCGCGCGCGCGCGCGCG 2232  
QY 547 CTCCTCTCTTGGGCTGCG 601  
Db 2233 AGCCCGCGCTGTGAGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2292  
QY 602 CCGGAGCGGAGCGGCG 661  
Db 2293 AGCTGCG 2352  
QY 662 GCGGCGCGGAGTGTGAGGAGAGATTCAGCGGCTCAAGCGCGCGCGCGCGCGCGCG 721  
Db 2353 TCGCGCGCGCGGAGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2412  
QY 722 GGGCGCTGCG-----GAGGCGCGGCGCGCTTTCGCGCGCGGCTGCGGAGAGCG 772  
Db 2413 GGGCG 2472  
QY 773 CCGACCGTCCGAGCG 832  
Db 2473 CCGCGGCGGAGCG 2532  
QY 833 GCGCGCGGAGACTTCGAGCGGCTTTCGCGGAGACTCCGCGCGCTCATGAGAGCGAG 892  
Db 2533 GCGCTACCG 2592  
QY 893 CCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 931  
Db 2593 CCGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2631

## RESULT 13

US-09-827-688-8  
; Sequence 8, Application US/09827688  
; Patent No. 6821955  
; GENERAL INFORMATION:  
; APPLICANT: ORSON, FRANK  
; APPLICANT: KINSEY, BERNIA  
; APPLICANT: BHOAGL, BALBIR  
; TITLE OF INVENTION: MACROMERGED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION  
; FILE REFERENCE: P01949US1/10004014  
; CURRENT FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: 60/195,680  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 8  
; LENGTH: 154746  
; TYPE: DNA  
; ORGANISM: HERPESVIRUS 2  
US-09-827-688-8

Query Match 10.6%; Score 102.6; DB 4; Length 154746;  
Best Local Similarity 46.8%; Pred. No. 4.8e-08;  
Matches 439; Conservative 0; Mismatches 484; Indels 16; Gaps 3;

QY 7 GGGTTCCTCCCGGAGCGTCTGCTGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 66  
Db 151404 GCG 151461  
QY 67 CTGCTCGGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 126  
Db 151462 GCGGTGCG 151521  
QY 127 CCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 186



[illegible]

```

; LENGTH: 154746
; TYPE: DNA
; ORGANISM: HERPESVIRUS 2
US-09-827-688-8

Query Match          10.6%; Score 102.6; DB 4; Length 154746;
Best Local Similarity 46.8%; Pred. No. 4,8e-08;
Matches 433; Conservative 0; Mismatches 481; Indels 16; Gaps 3;

QY 7 GGGCTTCCCCCGACGCTGCTGCTGCGGCGGACGCGGGGCGGTGGGCGGCGCATGTTGCGGGG 66
Db 130338 GCCCGCGCGCCGCGCGCGCCCGCTTGCGCTGCGGGCGGGCGGTGCGCGCGC--CGACGAGCGC 130281

QY 67 CTGCTCTGGGAGAGCGGGGACGCCGACGCTGCTGCTGCTGACCTCTGACCGCGCGCGGACCG 126
Db 130280 GCGGTGCTCCCGCGCGCTACGCGCGCGCGGGAGGTGCTGCGCGCGCTTGAGGCGCTTGAGCGCC 130221

QY 127 CCGGAGCGCTGCTGCTGCTGCTGCGACACTCAACCGCGCGGGGGCCGGAATCTGCGCGCGCCCTC 186
Db 130220 GCGCGCGCGCTCTCGCGCGCGCGCGCGGGGCGACACGACGACGACGACGCGCGCGCGGT 130161

QY 187 CCGGAGCGGAGACCTGCTGCTGCTGCTGCGCGTACGAGAGCGTGCGCTTCGAAAGCGGTGGCG 246
Db 130160 GGTGGCGGCGCGCGCGCGCGCGCGAGGCGGAGCGCGTGGCGCGTGGAGTGCCTGGCGGCTGC 130101

QY 247 CCGGTGACCCGCGCTCATGCGCGCGGGGCGCGCTGCTGCGCGACACTGTCCCTGCTCCGACG 306
Db 130100 CCGGAGATCTTGAGAGCGCTGCGGAGAGGAGGCTTTCAGCGGCGGCTGGCGCGCGTGC CGGG 130041

QY 307 GGCATGCGCGCGGAGAGCTCGCGGCGCCACGCGCGCGCGCGTGCAGACGTTGGGCGTCAACCG 366
Db 130040 CTGGCGCGAGCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129981

QY 367 ATGTTTGCCCCCGCGCGCGCGCATGACCGAGCGGACCCGTGCGCGCGCGTGTACAGGAGAC 426
Db 129980 CCGCAGCGGAGAGCGCGCGCGCGCGCTGCGCGCGCTGGCGCGAGCGAGTGCAGGTTCGTCGCGAC 129921

QY 427 GGGCGGGGCGGTACACGCGCGCTGCTGCGGCGTCTGTCAGAGGCGCGCGCGCGCGACAGGCCCGTACGG 486
Db 129920 GCGCTGATGCTATGCGCTGCGCGGAGGACCTGCGCGCGCGCGCGGAGAGAGAGCGCGC 129861

QY 487 CTCACGCGGAGAGAGACGACCGGCGGACGAGCGGCGGACCCAGGCGCTTGACGACGCGCGTG 546
Db 129860 GTGGCGCGCGTGGCGCGCGGTGAGCTGTGTCGCGCGGGGCGTGGGCGCGGCGGTGCGCGG 129801

QY 547 CTCTCTCTCTTGGGCGTGCCTCGCCCGCGCTCTCGGCGTTCGACGTCCGGGCGCTG----G 601
Db 129800 AGCGCGCGCGCTCTGAGCTGCCCGCGCGCGCGCGCGCGCGCGGACCTGCTCTTTCAGAACAG 129741

QY 602 CCGCGACGAGCACGCGCGCGCGCCCAACGAGGTGCTGCTGCGCGCTCTCTGAGGCGCGTGTGCTGGAG 661
Db 129740 AGCTTGGCGCGCTCTGCTGCGCGACACGTCGCGCGCGCGCGACACTGCTCGCGCGCGCGC 129681

QY 662 GCAGCGCCGAGGTTGATCGGAGACATTCAGCGGTCCAAACCCCGGAGCGAGCTTCGCGCGCC 721
Db 129680 TTCGCGCGCGGAGAGCGCGGCAAGGCAAGAGCCCGCGCGCGGACGAGGCGCGCGCGGCG 129621

QY 722 GGGCGCTCGC-----CGAGCGCTGCGCTCTCTTGGCGCGCTGCTGCTGCGGAGGAGAC 772
Db 129620 GGGCGCGCGCGCGCGCGCGCGGAGAGAGCGCGCGGACGCGCGCGCGCGCGCGCGCTT 129561

QY 773 CGGACCGGTGACGACGCGCGCGCGCGCGCGCGAGAGCGCGCGCGCGCATCCCGGGGAGATTGCGAG 832
Db 129560 CCGCGCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129501

QY 833 GCGCGCGGAACTTCAGCGCGCTTTCGAGGAACTTCGCGCGGCTTCATGGAGACGAGAGCTCG 892
Db 129500 GCGCTGACCGCGCGCGCGCGCGAGGAGCGCGGACCGCGCGCGGCGGTGAGCGCGGACGCG 129441

QY 893 CCGCGCGGACGAGACCACTTCGAGAGCTGTTCCGACCC 931
Db 129440 CCGGGGCGCGAGCACACGCGCGCGCGCTTGGCGCGCGCG 129402

```



```
RESULT 15
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patient No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
Query Match      10.4%; Score 100.4; DB 3; Length 4411529;
Best Local Similarity 45.3%; Pred. No. 7.7e-08;
Matches 405; Conservative 0; Mismatches 486; Indels 3; Gaps 1;
```

```
QY      2  TGAAGGCGCTTCCCGGAGCGTCTGTCGCGGACGCGGAGCGGCGGCGCATGTTTCG 61
Db      3934383  TGGCCGCGCTTGGCGCGCGGACACCACTTGCGCGCGCTTGAATGTTGTGCGCGGCGCCCG 3934324

QY      62  CCGGCGCTGTCGCGGAGGCGGCGGACGCGGACGCTGTCGTGACCTGTAACCGCGCGG 121
Db      3934323  CCAAGACCGCGCGCGCGCGCTTGGCCGCAATGGCCCGCGCGCGCGCGCGCTTG 3934264

QY      122  GACGCGCGGACGCGCTGCTGTGGGCGACGTACCGCGCGCGCGCGCGCGCGCGGAACTGCGCGCG 181
Db      3934263  GCTGTACAGGCGCGACGCGCTGTCGCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCACTGTC 3934204

QY      182  CCTTCGCGGAGCGGACGCTGTCGCTGTCGCGCGGACGAGACGTGCGCGCTCAAGGCGG 241
Db      3934203  CCGCGACCGCGCGCGCTTGTGCGCGAGACCTGCGGCACTGCGCGCGCTCCCGCGCTCG 3934144

QY      242  TGGCGCGCGTGAACCGCGCTATGCGGCGCGCGCGCTGTCGCGGACACCTGTCGCTCC 301
Db      3934143  CCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCTTGAATGCGCGCGCGCGCGCGCTGCGCG 3934084

QY      302  GAGCGGCGCATGCGCGCGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCA 361
Db      3934083  GCGCGCGCTTGGCGCGCGCTTGGCGCGCGCTTGAAGCGCGGAGCGCGCGCGCGCGGACGCTG 3934024

QY      362  ACCCGATGTTGCGCGCGCGCGCGCGCATGACCGGCGCGCGCGCGCGCGCGCTGATCA 421
Db      3934023  GCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTACCGCG 3933964

QY      422  GGGACGCGGCGCGCGCGTACGCGCGCTGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCG 481
Db      3933963  GCGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933904

QY      482  TACGGCTCAAGCGGAGAGACGACCGCGGACGACGCGGCGCGCGCGCGCGCGCGCGCG 541
Db      3933903  CCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933844

QY      542  CCGTGTCTCTCTCTTTCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 601
Db      3933843  CCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933787

QY      602  CCGCGACGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 661
Db      3933786  CCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933727

QY      662  GAGCGCGCGGAGTACGCGGAGCATCCAGCGGTCAACCGCGCGCGCGCGCGCGCGCG 721
```

```
Db      3933726  GCGCGCGTACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933667

QY      722  GAGCGCTGCGCGGAGCGCGCGCTGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 781
Db      3933666  GCGTGTGCTGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933607

QY      782  CCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 841
Db      3933606  CCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933547

QY      842  ACCTGACGCGCGTCTTTCGCGGAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 895
Db      3933546  CCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933493
```

```
Search completed: October 5, 2005, 10:06:57
Job time : 195.767 sec
```



This Page Blank (uspio)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 20:50:02 : Search time 951.313 Seconds  
(without alignments)  
7033.027 Million cell updates/sec

Title: US-10-089-514-5

Perfect score: 969

Sequence: 1 aagcgagcgtctcccccagag.....aagcgaggaagcagcatga 969

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7442561 seqs, 3452328358 residues

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/prodata/1/pubpna/PCR\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/prodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/prodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/prodata/1/pubpna/PCRUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/prodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/prodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/prodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/prodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 14: /cgn2\_6/prodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 17: /cgn2\_6/prodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 18: /cgn2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 19: /cgn2\_6/prodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 20: /cgn2\_6/prodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 21: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 22: /cgn2\_6/prodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 23: /cgn2\_6/prodata/1/pubpna/US11A\_PUBCOMB.seq:\*
- 24: /cgn2\_6/prodata/1/pubpna/US11\_NEW\_PUB.seq:\*
- 25: /cgn2\_6/prodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 26: /cgn2\_6/prodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	969	100.0	969	20	US-10-472-587-5
2	915.4	94.5	3305	14	US-10-126-927-68
3	913.8	94.3	12391	21	US-10-126-927-67
4	115.8	12.0	138203	21	US-10-819-386A-1
5	115.6	11.9	125401	19	US-10-203-295-35
6	114.8	11.8	65140	19	US-10-203-295-1
7	109.2	11.3	4569	15	US-10-156-761-7339

c	8	109.2	11.3	9025608	15	US-10-156-761-1	Sequence 1, Appl
c	9	108.6	11.2	1884	15	US-10-156-761-3440	Sequence 3440, Ap
c	10	107.4	11.0	969	20	US-10-472-587-5	Sequence 5, Appl
c	11	107	11.0	1707	15	US-10-156-761-4723	Sequence 4723, Ap
c	12	107	11.0	9025608	15	US-10-156-761-1	Sequence 1, Appl
c	13	106.2	11.0	3305	14	US-10-126-927-68	Sequence 68, Appl
c	14	105.4	10.9	12391	14	US-10-126-927-67	Sequence 67, Appl
c	15	105.2	10.9	1167	15	US-10-156-761-6049	Sequence 6049, Ap
c	16	104	10.7	13842	9	US-09-861-289-30	Sequence 30, Appl
c	17	104	10.7	13842	9	US-09-860-846-30	Sequence 30, Appl
c	18	104	10.7	13842	10	US-09-988-384B-30	Sequence 30, Appl
c	19	104	10.7	13842	10	US-09-836-821-30	Sequence 30, Appl
c	20	104	10.7	13842	16	US-10-271-889-30	Sequence 30, Appl
c	21	104	10.7	36778	9	US-09-861-289-5	Sequence 5, Appl
c	22	104	10.7	36778	9	US-09-860-846-5	Sequence 5, Appl
c	23	104	10.7	36778	10	US-09-836-821-5	Sequence 5, Appl
c	24	104	10.7	36778	16	US-10-271-889-48	Sequence 48, Appl
c	25	104	10.7	37948	10	US-09-988-384B-5	Sequence 5, Appl
c	26	104	10.7	38506	10	US-09-793-708-19	Sequence 19, Appl
c	27	104	10.7	38506	15	US-10-201-365-1	Sequence 1, Appl
c	28	104	10.7	38506	16	US-10-160-539-19	Sequence 19, Appl
c	29	104	10.7	38506	21	US-10-468-828-19	Sequence 19, Appl
c	30	102.6	10.6	3957	16	US-10-200-562-193	Sequence 193, Appl
c	31	102.6	10.6	3957	16	US-10-237-551-193	Sequence 193, App
c	32	102.6	10.6	3957	22	US-10-945-050-193	Sequence 8, Appl
c	33	102.6	10.6	154746	10	US-09-827-688-8	Sequence 8, Appl
c	34	102.6	10.6	154746	10	US-09-827-688-8	Sequence 8, Appl
c	35	101.8	10.5	3024	22	US-10-939-036-35	Sequence 35, Appl
c	36	101.4	10.5	1839	11	US-09-758-759-150	Sequence 150, App
c	37	101.4	10.5	5802	14	US-10-152-886-84	Sequence 84, Appl
c	38	101.4	10.5	5802	24	US-11-053-055-84	Sequence 84, Appl
c	39	101.4	10.5	5802	24	US-11-053-055-84	Sequence 84, Appl
c	40	101.4	10.5	15738	16	US-10-329-078-46	Sequence 46, Appl
c	41	101.4	10.5	61944	16	US-10-329-079-34	Sequence 34, Appl
c	42	101.4	10.5	109519	11	US-09-758-759-1	Sequence 1, Appl
c	43	100.4	10.4	2700	19	US-10-203-295-29	Sequence 29, Appl
c	44	100	10.3	88421	9	US-09-976-059-1	Sequence 1, Appl
c	45	99	10.2	2940	15	US-10-156-761-3788	Sequence 3788, Ap

#### ALIGNMENTS

RESULT 1

US-10-472-587-5

Sequence 5, Application US/10472587

Publication No. US20040214274A1

GENERAL INFORMATION:

APPLICANT: YANAI, Koji

APPLICANT: SUMIDA, Naomi

APPLICANT: WATANABE, Manabu

APPLICANT: MORIYA, Tetsuki

APPLICANT: MORAKAMI, Takechi

TITLE OF INVENTION: Transformants Producing Substance P1022 Derivatives, Methods for

TITLE OR INVENTION: Producing The Same And Novel Biosynthesis Genes

FILE REFERENCE: 2003-1302A/MMC/00144

CURRENT APPLICATION NUMBER: US/10/472,587

CURRENT FILING DATE: 2003-09-22

PRIOR APPLICATION NUMBER: 82227/2001

PRIOR FILING DATE: 2001-03-22

NUMBER OF SEQ ID NOS: 40

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 5

LENGTH: 969

TYPE: DNA

ORGANISM: Streptomyces venezuelae

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(966)

US-10-472-587-5

Query Match 100.0%; Score 969; DB 20; Length 969;

Best Local Similarity 100.0%; Pred. No. 3.4e-192;







Db	2547	GCCTGATCTCTCTCTTCTGGGGGTGCCTCGCCGCGCTCTGGAGTCAAGTCCGGGGCTTG	2606
Oy	601	GGCGGACGGGCAACCGCCGCGCCCAACGAGTGTCTCGCTCGCTCTTGAAGCCGTGTGCTCGG	660
Db	2607	GGGGGAGCGGCAACCGCCGCGCCCAACGAGTGTCTGTGCTCGCTCTTGAAGCCGTGTGCTCGG	2666
Oy	661	GGCAGCCCCGAGGTGTACGGGGACATCCAGCGGTCCAAACCCCGGGCGGGCTCCGCGGC	720
Db	2667	GGCAGCCCCGAGGTGTACGGGGACATCCAGCGGTCCAAACCCCGGGCGGGCTCCGCGGC	2726
Oy	721	CGGGCGCTCGCGGAGGCCCTGTGCGCTCTTTCGCGCGGTGTGTGGGAC-----	768
Db	2727	CGGGCGCTCGCGGAGGCCCTGTGCGCTCTTTCGCGCGGTGTGTGGGACGCCGAGACCG	2786
Oy	769	-----GACCCCGGACCGTGTCCGACGCGCCCCCGGGCGCGCCAGACGCCCCCGGCGCATTC	822
Db	2787	GCCGAGGACCCGAGCCGCGCGGACGCCACCCGACCGACCAACCCCGGCGCATTC	2846
Oy	823	GGATTCGAGCGGCGCGCGGGAACCTCGACGCGCTTTTCGGGGAACCTCCCGGCTCATGGGA	882
Db	2847	GGATTCGAGCGGCGCGCGGGAACCTCGACGCGCTTTTCGGGGAACCTCCCGGCTCATGGGA	2906
Oy	883	CCGAGCTCGGGGGGGGCGCAGAGCAACATGCGCAGGAGCTTCCGACCCCTCCACCGCAC	942
Db	2907	CCGAGCTCGGGGGGGGCGCAGAGCAACATGCGCAGGAGCTTCCGACCCCTCCACCGCAC	2966
Oy	943	GACGACGAAGCGAGAACGACCGATGA	969
Db	2967	GACGACGAAGCGAGAACGACCGATGA	2993

### RESULT 3

```

Sequence 67, Application US//10126927
Publication No. US20030082575A1
GENERAL INFORMATION:
APPLICANT: The Scripps Research Institute
APPLICANT: Schultz, Peter G
APPLICANT: Wang, Lei
APPLICANT: Anderson, John C
APPLICANT: Chin, Jason
APPLICANT: Liu, David R
APPLICANT: Magliery, Thomas
APPLICANT: Messgers, Eric L
APPLICANT: Mehl, Ryan A
APPLICANT: Pasternak, Mirol
APPLICANT: Santoro, Stephen W
APPLICANT: Zhang, Zhilwen
TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
FILE REFERENCE: 54-000120US
CURRENT APPLICATION NUMBER: US/10/126,927
CURRENT FILING DATE: 2002-04-19
PRIOR APPLICATION NUMBER: US 60/285,030
PRIOR FILING DATE: 2001-04-19
PRIOR APPLICATION NUMBER: US 60/355,514
PRIOR FILING DATE: 2002-02-06
NUMBER OF SEQ ID NOS: 79
SOFTWARE: Patentin version 3.1
SEQ ID NO 67
LENGTH: 12391
TYPE: DNA
ORGANISM: Plasmid pSC101, Streptomyces venezuelae pAPABC
US-10-126-927-67

```

```

Query Match 94.3% Score 913.8: DB 14 Length 12391;
Best Local Similarity 96.5% Pred. NO. 4.3e-181;
Matches 952; Conservative 0; Mismatches 17; Indels 18; Gaps 1;

OY 1 ATGAGCGGCTTCCCCGACAGCTGTCTCTGGCGGCAACGGGGCGCTGTGGCGGCATGTTCC 60
    |||||
Db 2361 ATGATGGCTTCCCCCGAGGCTCTCTCTCGCGGCAACGGAGCGGTGGCGGCATGTTCC 2420

OY 61 GCCGGGCTGTCGGGAGAGCGGAGCGGACGCTCTGTGTGACCTTGATCGCGCGCGG 120

```

[illegible]

## RESULT 4

US-10-819-386A-1/c  
Sequence 1, Application US/10819386A  
Publication No. US20050089884A1  
GENERAL INFORMATION  
APPLICANT: Korea advanced Institute of Science and Technology  
APPLICANT: Shanghai Jiaotong University  
APPLICANT: LEE, Sang Yup  
APPLICANT: DENG, ZIXIN







```

Db 28857 TCGCGGTCGCGCGCACCGGCGTCAAGCCCGTGCCTCGTCCGTAAGCCGCGCGCGCG 28916
QY 383 CCGGCAATGACCGGCGGACCGCTGCGCGCGCTGCTACACAGGAGACGGCGCGGCTACGG 442
Db 28917 ACCGTGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 28976
QY 443 CCGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 487
Db 28977 GCAACATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29036
QY 488 TCACGCGCGAGAGACGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 547
Db 29037 TGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29093
QY 548 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 607
Db 29094 TGGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 29153
QY 608 CGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 667
Db 29154 TCGCGCGGCTCTCTGACGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29213
QY 668 CCGAGGTGTAACGAGGACATCAGCGGTCAACCCCGCGCGCGCGCGCGCGCGCGCGCG 727
Db 29214 CGGAGGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29273
QY 728 TCGCGGAGCGCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 787
Db 29274 CCGCGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 29333
QY 788 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 847
Db 29334 TCGCGGCGCTCTGAGGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 29393
QY 848 ACCGCGCTCTGCGGGAACCTCCGCGGCTCATGGAGCTGCGCGCGCGCGCGCGCGCG 907
Db 29394 ACCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29453
QY 908 ACTGCCAGAGCTGTTCGCGACCCCTCCAGCGCACCGAC 945
Db 29454 CCGGCGCGGACCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 29491

```

```

; SEQ ID NO 1
; LENGTH: 65140
; TYPE: DNA
; ORGANISM: Streptomyces noursei ATCC 11455
US-10-203-295-1
Query Match 11.8%; Score 114.8; DB 19; Length 65140;
Best Local Similarity 46.9%; Pred. No. 1e-15;
Matches 449; Conservative 0; Mismatches 487; Indels 22; Gaps 2;

QY 27 CGTGCGCGGACGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCG 86
Db 19867 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19926
QY 87 CCGCAAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 146
Db 19927 AGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19986
QY 147 CGAGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 206
Db 19987 CGAGGACACCGTACCATGCGCGCTTACGACGCGCGCGCGCGCGCGCGCGCGCG 20046
QY 207 GCTGCGCGTACAGAGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 266
Db 20047 CTCCTGTGCTCCCGGAGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20106
QY 267 GCGGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 326
Db 20107 CGACTCTCTTTCAGAGTGAAGTGAACCCGCTCTCAAGGCGCGCGCGCGCGCGCG 20166
QY 327 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 386
Db 20167 CACTGTGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20226
QY 387 CATGACCGGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 446
Db 20227 CATTGCGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20286
QY 447 GCTGCGGCTGCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 506
Db 20287 CGACTGTGCTGACCACTTCAACACACCGCGCGCGCGCGCGCGCGCGCGCGCG 20346
QY 507 CCGGAGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 566
Db 20347 CGCACAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20406
QY 567 CTTGCGCGCGCTGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 626
Db 20407 CGACGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20464
QY 627 GGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 686
Db 20465 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20524
QY 667 CCGGAGGTGTAAGGAGACATCCAGCGGTCCAAACCCCGCGCGCGCGCGCGCGCG 726
Db 20525 CTGACCTTGCCTCCCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20584
QY 727 CTGCGGAGGCGCTGCGCTCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 786
Db 20585 GCCAGGCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20644
QY 787 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 846
Db 20645 GCCGCTGCTCCGCTCCCGACCGAACCACTGGAACCGGACCGGACCTGCTGATC 20704
QY 847 GAGCGGTGTTGAGGAGACTCGCGCGCTCATAGGACCGAGGTGCGCGCGCGCGAGAC 906
Db 20705 ACCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20764
QY 907 CACTGCCAGAGCTGTTCCGCACTTCACCGGACCGAGCGAGCGAGGAGGAGGAGGAGGAG 964
Db 20765 GTCCGCACTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20822

```







Db 8781068 GTCGGGCCCAACCGGAAACGCTGCGCTGCGCCCGCAAGCGCTGCGGAAAGGCGACGG 8781009  
QY 217 CACGAGACGCTGCGCTCAAGGCGCTGCGCGCCCGTACCGCGCTCATGCGCGCGCGCG 276  
Db 8781008 TGGCGGGGCGCGGCGGACGCTGCTGCTGCTGCGAGCCCGCGCTGCGGCGCGCGCG 8780949  
QY 277 CTGCTCGCGCAACACCTGCTGCTGCTGCGAGCGGCACTGCGCGG---AGCTGCGCGCGCAC 333  
Db 8780948 AAGCTGCTCTTCTGCTTCCCGCGCTGAGGCGGAGTTCTGCGCCAACTCGGAGAGCTC 8780889  
QY 334 GCGCGCGCGCTGCGACGCTGCGGCTCAACCGGATGTTGCGCGCGCGCGCGCGCATACC 393  
Db 8780888 GCGCGGACCTTCTGCGGCTGCGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCATGCT 8780829  
QY 394 GCGCGACCGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 453  
Db 8780828 GCGCGGAGGCGACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8780769  
QY 454 CTGCTCGAGGCGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 513  
Db 8780768 CTGCGGCTGCGACCG 8780709  
QY 514 ACAGCGCGCAACCGAGCGCTGACGACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 573  
Db 8780708 GCGCGCGAAATCCACGACCGGACGAGGTGACGCGCTTCTGCGCGCGCGCGCGCG 8780649  
QY 574 GCGCTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 633  
Db 8780648 GCGCTGAGCGCTTCCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8780589  
QY 634 CTGCGCTGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 693  
Db 8780588 ACAGAGTGGCG 8780529  
QY 694 TCCAAACCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 753  
Db 8780528 ATGATCTGCG 8780469  
QY 754 GCGCTGCTGCG 813  
Db 8780468 CTGATCTGCGAGATCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8780409  
QY 814 CATCCGCGGAGATGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 873  
Db 8780408 CTGCGCGCGATCGAGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8780349  
QY 874 CTGATGCGAGCGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 933  
Db 8780348 TGGTCGCGACGACCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8780289  
QY 934 CACCGC 939  
Db 8780288 GTCGCG 8780283

## RESULT 9

US-10-156-761-3440  
; Sequence 3440, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 3440  
; LENGTH: 1884  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1884)  
US-10-156-761-3440  
Query Match 11.2%; Score 108.6; DB 15; Length 1884;  
Best Local Similarity 45.5%; Pred. No. 5,9e-14;  
Matches 436; Conservative 0; Mismatches 514; Indels 9; Gaps 1;  
QY 6 CCGCTTCCCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 65  
Db 855 CGCGTCCCGGACCTTCAACGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 914  
QY 66 GCTGCTGCGGAGCGGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 125  
Db 915 GGTCTTGGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 974  
QY 126 GCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 185  
Db 975 CCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1034  
QY 186 CCGGAGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 245  
Db 1035 CCGGAGTCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1094  
QY 246 GCGCTGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 305  
Db 1095 GGTGGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1154  
QY 306 GCGCATGCGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 365  
Db 1155 CGCGTTCACGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1214  
QY 366 GATGTTGCGCCCGG-----CGCGCGCATGACCGGCGGACCGCGGCGCGCGCTGCTGCT 416  
Db 1215 GAGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1274  
QY 417 CACCAAGGACGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 476  
Db 1275 CGCGGCGGAGCGCGCGGAGCGGACCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1334  
QY 477 GCGCTGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 536  
Db 1335 GCGCTTGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1394  
QY 537 GCAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 596  
Db 1395 CGCGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1454  
QY 597 CTGCGGCGGACG 656  
Db 1455 CGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1514  
QY 657 CGCGGCGACCGCGGAGTGAAGGAGCATCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 716  
Db 1515 CGTGTGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1574  
QY 717 GCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 776  
Db 1575 GAGCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1634  
QY 777 CCGTCCGACG 836  
Db 1635 CCGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1694  
QY 837 CCGGAGCTGACGCGGCTGCTGCGGAGACTCGCGCGGCTCATGAGACCGGAGCTCGCGCG 896



```
Db      1695 CGACCCCTTCGCGCGGCGCTCGTCGCGCCAGGTGCGGCCCTTCCGCGCAGAGACTGACCGG 1754
QY      897 GGCGACGAGACCACTGCGAGAGGCTGTTCCGACACCTTCACCGACCGAGAGAGAGCGG 955
Db      1755 GACCTGTGAGTCCGCGGCGCGCTCGCTCGCGCGCAGCGCGTGAACCGCGCTCCAGAGCGG 1813

RESULT 10
US-10-472-587-5/c
; Sequence 5, Application US/10472587
; Publication No. US20040214274A1
; GENERAL INFORMATION:
; APPLICANT: YANAI, KOJI
; APPLICANT: SUMIDA, Naomi
; APPLICANT: WATANABE, Manabu
; APPLICANT: MORIYA, Tatsuki
; APPLICANT: MURAKAMI, Takeshi
; TITLE OF INVENTION: Transformants Producing Substance Pf1022 Derivatives, Methods for
; FILE REFERENCE: 2003-1302A/MMC/00144
; CURRENT APPLICATION NUMBER: US/10/472,587
; CURRENT FILING DATE: 2003-09-22
; PRIOR APPLICATION NUMBER: 82227/2001
; PRIOR FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 5
; LENGTH: 969
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(966)
US-10-472-587-5

Query Match      11.1%; Score 107.4; DB 20; Length 969;
Best Local Similarity 47.7%; Pred. No. 1.3e-13;
Matches 378; Conservative 0; Mismatches 411; Indels 4; Gaps 2;
```

```
Db      333 GTGGGCGCGAGACTCCGCGGCGCATGCGCTCCGAGAGGACAGAGGTGTCCGCGAGAGCGCC 274
QY      561 GCTTCGCGCTTCGCGCGCGCTGCGGTGACGTCCGCGCGCTTGGGCGCGACCGCGCGCG 620
Db      273 GCGCGCGCGCATGAGCGCGGCTGACCGGCGCGCACCGGCTTGAAGGCCACGCTCTCTGTGTAAC 214
QY      621 --CAACAGATGCTGCTCGCGCTCTCTGCGCGCGCTGCTGCTGCGCGCGCGCGCGCGCGCGCG 678
Db      213 GCGAGACAGAGAGAGAGTCCGCGCTCCCGAGAGGCGCGCGGAGTTCGCGCGCGCGCG 154
QY      679 GGGAGATTCAGCGGTTCACACCCCGGCGCGCGGTCCGCGCGCGCGCGCTTCGCGAGGCC 738
Db      153 GAGTTCGCGCACACGAGCAGGCGCTCCGCGCGCGCGCGCGCGCGCGGTACGAGGTTCAGACGAG 94
QY      739 CTGCGCTCTTGGCGCGCGCTGAGTCCGCGCGAGACCGCGAGCCGTCGCGCGCGCGCGCGCG 798
Db      93 CTGCGCGCTGCGCGCGCTCTCCGAGAGACCGCGCGAGAACATGCGCGCCACCGCGCGCTGCC 34
QY      799 GCCGACGCGCGCG 811
Db      33 GCGGACGAGCAGCG 21

RESULT 11
US-10-156-761-4723
; Sequence 4723, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 4723
; LENGTH: 1707
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1707)
US-10-156-761-4723

Query Match      11.0%; Score 107; DB 15; Length 1707;
Best Local Similarity 48.0%; Pred. No. 1.3e-13;
Matches 342; Conservative 0; Mismatches 365; Indels 6; Gaps 1;
```



Oy 478 CCGGTACGGCTTCACGGCGGAGGAGGACGACCGGAGCGAGGCGGACCCAGAGCCCTTAGC 537  
 Db 1060 AGCCGCGCACGCGCCGAGGTGAGGACCGAGCTGATACCGGGCTGCGGCCACGGTGGCC 1119  
 Oy 538 CACGCGGTGCTCTTCTCTTGGGGCTGGCCCTTGGCCGCGCTTGCGGGTGAAGTCCGGAGC 597  
 Db 1120 GCCGCGCGCGCTCTCCCGAGGCGCGCGCCGAGCCCGTGTGAGAGGCCCGCCCGAGCC 1179  
 Oy 598 CTGGCGGGGACGCGACCGCGCCGCCACACAGGTGTGCTGCGCCCTCTGAGCCCGGTGTCTC 657  
 Db 1180 GCGCGCGGCTTCGCACTGTGCACTTCCGAGTCCCGTGTGACATGACGCGCGCGTGTAC 1239  
 Oy 658 GCGCGACGCCCGAGGTATACGGGACATTCAGCGGTCAACCCCGCGCGCGGTCCGCG 717  
 Db 1240 CCGGTCTCTGGGGGTGGCGGCTGTGACATCGAGTGTG-----CTGCGCGAGGGCTTCGCG 1293  
 Oy 718 CGCCGGGCGCTTCGCGGAGGCCCTTGTGCTCTTCCGCGCGTGTGTGCGACGACCCGGAG 777  
 Db 1294 GACGCGCGCGATGGCGCGCGCGAGGTGACCTGTGTGCGCCCGGATCTGTGCGCGAGAGGTGGCC 1353  
 Oy 778 CGTGGCGACGCGCCCCGGGCGCGCGACGCCGCCCGGCATTCGCGGGGAGATGGACGCGCGCC 837  
 Db 1354 GCCTTTCGGCGCGCGCCAGCGCGCGCGCGACATCACCCACCGTGTGTGCTTGGCGACG 1413  
 Oy 838 GGGAACTTCGACCGCGCTTTCGGGGAACTTCGCGCGGCTCATGGAGCCGGAAGCTTCGCGCG 897  
 Db 1414 ATGGCGCGCGAGTGTGTGCGACGAGATTCGGCGGGCTTCGACGAGCGGCTGCCGGGCTTC 1473  
 Oy 898 GGCACGAGACCACTGCGAGAGCTGTTCGCGACCTTCACCGCACCGACGAGAGAGA 950  
 Db 1474 GACGAGAAAGCAGCGCGCGAGATCACCCGAGACCGTGGCGCGGTGTGTGACAA 1526

```

RESULT 12
US-10-156-761-1
Sequence 1, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
PRIORITY FILING DATE: 2002-05-29
PRIORITY APPLICATION NUMBER: JP 2001-204089
PRIORITY FILING DATE: 2001-05-30
PRIORITY APPLICATION NUMBER: JP 2001-272697
PRIORITY FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

```

Query Match	11.0%	Score 10.7	DB 15	Length 9025608
Best Local Similarity	48.0%	Pred. No. 9.5e-15		
Matches 342	Conservative 0	Mismatches 365	Indels 6	Gaps 1
QY	238	GCCGTGGCGCCCGGTACCCGGGCTCATGCGGCGCGCGCGCTGCTCCGCGACACCTGTTC	297	
Db	5778628	GCGCGGGCGGTGTGAGGGGCGTACGGGTACCGCCGGTGGCGGTGCGGTGAGGAGAACCCCGGC	5778668	

OY	298	TTCCGAGACGGGCAATGCGCGCGAGAGCTGCGGAGCCACGACCGCCCGGCGTCCAGAGAGTGGG	357
Db	5778688	TTCCGGGCGGGCGGGCTTCGCCCCCGCGGATACCGAGAGGAGCTGCCCTTCGACTGTCC	5778747
OY	358	CTCAACCGATGTTTCGCCCCCGCGCGCGGAGATGACCGGCGGACCCGTGGCGCGTGGTCTC	417
Db	5778748	GCGGTTCAGGGGCGGCCAGGGGCTTCTCCGTATGGGAGAGGACCGCGGATCGCCGCGGATGGAC	5778807
OY	418	ACCAAGGAGACGGGAGCGGGGCTTCAGGCCCCCTGCGGTGCGGTCTCCAGGGCGCGCGGGCAGG	477
Db	5778808	GCGGCGAACCCCTGGAAACAGCAGCGCGGCTCTGGAGTGCACCGGAGAACCTGTGCAGCGCGGAC	5778867
OY	478	CCCGTACCGGCTCAACGCGCGGAGAGACGACCGGAGCGCGGACCGACCCAGGCGCTTAGC	537
Db	5778868	AGCGCGCGGCAACGCCCGGAGTTCAGAGCCGAGCTATACCGGCGTCCGCGCACGAGTGGC	5778927
OY	538	CACGCGGTGCTCTCTCTTCGAGGCTGCGCCCTCGCGCCGCTCGGCGTGCAGCTCCGAGC	597
Db	5778928	GCCGCGGCGGCGTCTCCCGAGCGCGCGCGGCGCCGAGCGCTCGTCGAGGCGCCCCCGGCC	5778987
OY	598	CTGGCGGCGGAGACGAGCGCGCGGCCCGCCACAGATGCTGCTCCGCCCTCAGGCGCGGTGCTC	657
Db	5778988	GCGCGCGCGCTCGCACTGCTTCGACTTCGATGCGATGCGCCCTGTGATCAGCGCCCGTGCAC	5779047
OY	658	GCGCGGACGCCCGGAGTGTACGGGGACATCAGAGGTCCAAACCCCGGCGCGGTCCGCG	717
Db	5779048	CGGCTGCTCGGGGGTGGCGGCTGTGTCATCGAGTGC-----CTCGCGAGAGGCTTCGCG	5779107
OY	718	CGCGCGGCGCTTCGCGGAGCGCTTCGCTCTTCGCGCGCGCTGGTTCGGCGAGACCGCGGAC	777
Db	5779102	GACCGCGCGATGCGCGCGCGGACGTGAGACTTGAGCGCGATGCTCGCGCGAGAGTGGCC	5779161
OY	778	CGTCCGACGCGCCCGCGGCGCGCGACGCGCCCGGCATCCCGGGGAGATCGACGCGGCG	837
Db	5779162	GCTTTCGGCGCGCGCGACGCGCGCGCGGACATACCCCGACCGTGTGTGCGCTTCGGCAGCG	5779221
OY	838	GCGAACCTCGACGCGGCTTTCGGGGAACTCCGCGGCTCATGAGGACCGAGACTTCGCGAGC	897
Db	5779222	ATGGCGCCCGACGTGTGTGGGAAAGATGACGCGGCTTCGACGGAACGCGCTGCGCGGCTC	5779281
OY	898	GCGCAGAGACATCGCCGACGAGCTGTTCCGACCTCCACCGGACCGGACGAGCGA	950
Db	5779282	GACGAGAGACGCGCGCGCGAGATACCCAGACCGTGGCGCGCTGTGTGCACAA	5779334

```

; RESULT 13
; US-10-126-927-68/C
; Sequence 68, Application US/10126927
; Publication No. US20030082575A1
; GENERAL INFORMATION:
; APPLICANT: The Scripps Research Institute
; APPLICANT: Schultz, Peter G
; APPLICANT: Wang, Lei
; APPLICANT: Anderson, John C
; APPLICANT: Chin, Jason
; APPLICANT: Liu, David R
; APPLICANT: Megliery, Thomas
; APPLICANT: Meggers, Eric L
; APPLICANT: Mehl, Ryan A
; APPLICANT: Pastinak, Miro
; APPLICANT: Santoro, Stephen W
; APPLICANT: Zhang, Zhiwen
; TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
; FILE REFERENCE: 54-000120US
; CURRENT APPLICATION NUMBER: US/10/126,927
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,030
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 60/355,514
; PRIOR FILING DATE: 2002-02-06
; NUMBER OF SEQ. ID NOS: 79
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 68

```



```
; LENGTH: 3305
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-10-126-927-68
```

Query Match	11.0%;	Score 106.2;	DB 14;	Length 3305;
Best Local Similarity	48.2%;	Pred. No. 1.6e-13;		
Matches 391;	Conservative 0;	Mismatches 413;	Indels 7;	Gaps 3;

[illegible]

RESULT 14  
US-10-126-927-67/c

Sequence 67, Application US/10126527  
Publication No. US20030082575A1  
GENERAL INFORMATION:  
APPLICANT: The Scripps Research Institute  
APPLICANT: Schultz, Peter G

```

? APPLICANT: Wang, Lei
? APPLICANT: Anderson, John C
? APPLICANT: Chin, Jason
? APPLICANT: Liu, David R
? APPLICANT: Magliery, Thomas
? APPLICANT: Meggers, Eric L
? APPLICANT: Mehl, Ryan A
? APPLICANT: Pasternak, Miro
? APPLICANT: Santoro, Stephen W
? APPLICANT: Zhang, Zhilwen
? TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
? FILE REFERENCE: 54-000120US
? CURRENT APPLICATION NUMBER: US/10/126,927
? CURRENT FILING DATE: 2002-04-19
? PRIOR APPLICATION NUMBER: US 60/285,030
? PRIOR FILING DATE: 2001-04-19
? PRIOR APPLICATION NUMBER: US 60/355,514
? PRIOR FILING DATE: 2002-02-06
? NUMBER OF SEQ ID NOS: 79
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 67
? LENGTH: 12391
? TYPE: DNA
? ORGANISM: Plasmid pSC101, Streptomyces venezuelae papABC
? US-10-126-927-67

```

Query Match	10.9%;	Score 105.4;	DB 14,	Length 12391;
Best Local Similarity	48.2%;	Pred. No. 1.5e-13;		
Matches 389; Conservative	0;	Mismatches 411;	Indels 7;	Gaps 3;

QY	21	CGTCGTGATCGAGCGGACGAGCGGCGGAGGCGGACATTTCCGCGGAGCTGCTCGGAGAGGC	80
Db	3171	CGGGGTGTGTGGCGCGGATCCGGGTCTCTCGGCGCGGTCCGGGTGTGTCCGCATCAGCGCG	3112
QY	81	GGGCAAGCCGACGCTGTGTGTGACTCTGTACCGCGCGCGGACGAGCTGCCTGCT	140
Db	3111	CGAAGGAGCGCAGAGGCTCTCGGCGAGCGCCGCGCGCGAGACGCGCGCCGGGGGTTTGAAC	3052
QY	141	GTATGGGGAAGTACCGCGCGCGGAGCCCGAACTGTGCGCGCGCTCTCCGGGACGCGGACT	200
Db	3051	GCTGTGATGTCCCGGTACACTTCCGAGGTGCGCGCGACACACGAGGCCAGGAGGCGAGCA	2992
QY	201	CGTCTGCTTGCACGTAACACGAGACGTGAGGCTCTCAAGGCGGTGAGCGCCGTGACCCGACT	260
Db	2991	GCACTGTGTGG--GGCGGCGGTGCGGTGCGCGCGCAGGAGCCGAGACGTGACGCCGAGAGCG	2934
QY	261	CATGCGGCGGGGCGCGCTGCTCGCGCACACCTGTCTCGTCCGACGCGGACATGCGCGGA	320
Db	2933	GGGAGGCGCAGGCGCCGAGGAGAGGATCACGAGCGTGTGACAGGCGCTTGTGGTCCGCGCGT	2874
QY	321	GCTCGGAGGCCACAGCCCGCGCGGTACGACAGTGTGGGCTCAACCCGATGTTTGGCCCCGC	380
Db	2873	CGTCCGCTCTGCTCTCTCCCGCTGAGCCGTACGAGGCGTGTGCGCGCCGCTTGACGAG	2814
QY	381	CGCGGCAATGACCGGCGGACCGGTGCGCGCGCTGTGTACACAGGAGCGAGCGCGGCGTAC	440
Db	2813	CCGACAGAGGCGGTGACGCGCGGCGCGTCCGTGTGACACAGCGCGGCGCACAGGCGGCG	2754
QY	441	GGCCCTCTGTGCGGCTGTGAGGAGCGGCGGCGGCGCGCTTACAGCGCGGAGGA	500
Db	2753	GGTCAATGCGCGGCGGCGGCGGAGCAATCGGGTTTGTAGGCCCAACGTGCTGTGAGACGCGCGGAGGC	2694
QY	501	GCACGACCCGGAACGACGCGCGGCGCAACCAAGGCGCTGTGACGACAGCGGTGTCTTCCTTCG	560
Db	2693	GTGAGCGCGAGCTCCGCGGCGCATGCGCGTCCGACCGAACAGGATGTGCGCGAGACAGGC	2634
QY	561	GCTGCGCCTGACCAGCTCGGAGCTGTGACGTCTCGGAGCGCTTGTGCGCGACGAGCACGCGGC	620
Db	2633	GCCCGGTGTGATGACCGGAGTCAAGGCGCGCACAGGCTTGTAGGGCCACGTCTGTGTAC	2574
QY	621	--CCACCAAGTGTCTGTGCGCCCTCTGTGAGCCCGTGTGTGAGCGGACAGCCCGAGGTATC	678
Db	2573	GGGACACAGACGAGGTGCGCGTCCGAGAGGAGCGCGCGCAGCTCGGAGCCCGCGCGGT	2514







This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 20:20:08 ; Search time 3664.34 Seconds  
(without alignments)  
10065.737 Million cell updates/sec

Title: US-10-089-514-5

Perfect score: 969  
1 atgcgcgcgcctcccccgcag.....aagcgagagaagcagcatga 969

Sequence: IDENTITY\_NUC

Scoring table: Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : Listing first 45 summaries

EST: \*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	134.6	13.9	1462	AG441877	Mus muscu
C 2	133	13.7	1469	AG365356	Mus muscu
C 3	131.2	13.5	1610	B2569386	pacs2-164
C 4	128.6	13.3	1406	CG756569	P051-4-B0
C 5	127	13.1	2243	AG381986	Mus muscu
C 6	126.8	13.1	1137	BG809979	mgct002xd
C 7	126.8	13.1	1350	BM450237	LOCUS
C 8	126.6	13.1	1780	AG448243	Mus muscu
C 9	126	13.0	1198	CC190008	CH261-138
C 10	125.8	13.0	1692	AG396765	Mus muscu
C 11	125.2	12.9	1073	CK209412	FGAS02117
C 12	125.2	12.9	1821	CL090560	CH216-160
C 13	124.6	12.9	1218	CL081011	CH216-160
C 14	124.6	12.9	1267	CL476590	SATL_258
C 15	124.4	12.8	2332	AG363333	Mus muscu
C 16	123.8	12.8	1355	AG429718	Mus muscu
C 17	123.6	12.8	1569	AG341503	Mus muscu
C 18	123.6	12.8	2332	AG363333	Mus muscu
C 19	123.4	12.7	1878	AG393083	Mus muscu
C 20	123.2	12.7	1452	AG032979	Pan trogl
C 21	123	12.7	982	BQ687717	AGENCOURT
C 22	123	12.7	1281	BG852363	1024034A0
C 23	122.8	12.7	1070	B2557810	pacs1-60
C 24	122.8	12.7	1359	CL507830	SATL_788

25	122.8	12.7	1421	8	B2569488	pacs2-164
26	122.4	12.6	1280	9	CL470317	SATL_140
C 27	122	12.6	2077	9	AG435283	Mus muscu
C 28	121.4	12.5	1280	9	CL470317	SATL_140
C 29	121.2	12.5	1956	9	CG754548	P050-1-A1
C 30	120.8	12.5	1885	9	AG363563	Mus muscu
C 31	120.8	12.5	1628	9	CG757066	P052-2-A0
C 32	120.8	12.5	1822	9	AG435170	Mus muscu
C 33	120.6	12.4	1216	9	AG441574	Mus muscu
C 34	120.6	12.4	1300	9	AG430756	Mus muscu
C 35	120.4	12.4	1131	9	AG042920	Pan trogl
C 36	120.4	12.4	1552	9	AG430101	Mus muscu
C 37	120.2	12.4	1339	4	BM458211	AGENCOURT
C 38	120.2	12.4	1853	9	AG382854	Mus muscu
C 39	119.6	12.3	1625	9	AG043477	Pan trogl
C 40	119.4	12.3	1116	8	B2569478	pacs2-164
C 41	119.4	12.3	1360	9	AG136785	Pan trogl
C 42	119	12.3	1360	9	CL469817	SATL_133
C 43	118.8	12.3	885	8	AG159162	Pan trogl
C 44	118.8	12.3	965	8	AO893417	HS 4832 A
C 45	118.6	12.2	1674	9	CL078342	CH216-149

## ALIGNMENTS

RESULT 1  
AG441877/c  
LOCUS  
DEFINITION  
Mus musculus molossinus DNA, clone:MSMG01-323H09.TJ, genomic survey  
sequence.  
ACCESSION  
AG441877  
VERSION  
AG441877.1  
GI:48084940  
KEYWORDS  
GSS.  
SOURCE  
Mus musculus molossinus  
ORGANISM  
Mus musculus molossinus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.  
BAC end Sequences of Library MSMG01  
TITLE  
Unpublished  
JOURNAL  
2 (bases 1 to 1462)  
AUTHORS  
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.  
REFERENCE  
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suenho-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
COMMENT  
Clones are derived from the mouse BAC library MSMG01. For BAC  
library availability, please contact Kunya Abe (abe@rtc.riken.jp).  
Tsukuba Institute, Bio Resource Center,  
The Institute of Physical and Chemical Research (RIKEN) 3-1-1  
Koyada, Tsukuba, 305-0074 Japan  
Phone: 81-298-36-9189, Fax: 81-298-36-9199  
e-mail: abe@rtc.riken.jp  
PRIMERS  
Sequencing : TJ  
LIBRARY  
Vector : pBAC3.6  
R Site 1 : EcoRI  
R Site 2 : EcoRI.  
Location/Qualifiers  
1. 1462  
/organism="Mus musculus molossinus"  
/mol\_type="genomic DNA"  
/sub\_species="molossinus"  
/db\_xref="taxon:57486"  
/clone="MSMG01-323H09.TJ"  
/sex="male"  
/issue\_type="mixture of kidney and spleen"  
/clone\_id="MSMG01 Mouse Male BAC library"



## ORIGIN

Query Match 13.9%; Score 134.6; DB 9; Length 1462;

Best Local Similarity 50.7%; Pred. No. 6e-13;

Matches 442; Conservative 0; Mismatches 396; Indels 14; Gaps 4;

```
QY 5 GCGGCTTTCCCGGACGCTGCTGTCGCGGACGCGGGGCGGTGGCGGATTTGCGCG 64
    |||||
DB 1352 GCGGGCGCGCGCGCGCGCGCGGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1293
QY 65 GCGTGTGCGGGAGGCGGGGAGCGGACGCTGCTGTCGACCTTCGTAACCGCCCGGAG 124
    |||||
DB 1292 GCG-GCGGGGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1234
QY 125 GCGCGGACGCGCTGCTGTCGCGGACGCTGTCGCGCGCGCGCGCGCGCGCGCGCG 184
    |||||
DB 1233 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1175
QY 185 TCGGGAGCGGACCTGCTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 244
    |||||
DB 1174 GCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1115
QY 245 CGCGCGTGAACCGCGCTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 304
    |||||
DB 1114 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1055
QY 305 CGGGGATGCGCGCGGAGCTGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 364
    |||||
DB 1054 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 995
QY 365 CGATGTTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 414
    |||||
DB 994 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 935
QY 415 GTACACGAGGAGCGGGCGCGCGCGCTGACGCGCGCTGCTGCTGCTGCTGCTGCTG 474
    |||||
DB 934 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 875
QY 475 AGCGCCGTAAGCTCAACGCGGAGAGACGACGCGGACGCGGCGCGCGCGCGCGCG 534
    |||||
DB 874 CGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 815
QY 535 AGCGACGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 594
    |||||
DB 814 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 755
QY 595 GCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 654
    |||||
DB 754 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 695
QY 655 CTGCGCGGAGCGCGCGGAGTGTACGCGGAGATCCAGCGGTCGAACCGCGCGGCGG 714
    |||||
DB 694 CGCGGGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 635
QY 715 GCGCGCGCGCGCGCTGCGCGGAGCGCTG--CGCTCTTTCGCGCGCGCTGTCGCG 772
    |||||
DB 634 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 575
QY 773 CGGACCGCTGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 824
    |||||
DB 574 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 523
```

## RESULT 2

AG365356/c

LOCUS AG365356 1469 bp DNA linear GSS 03-JUN-2004

DEFINITION Mus musculus molossinus DNA, clone:MSMg01-168A20.T7, genomic survey

ACCESSION AG365356

VERSION AG365356.1 GI:47976561

KEYWORDS GSS.

SOURCE Mus musculus molossinus

ORGANISM Mus musculus molossinus

## REFERENCE

AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.

TITLE BAC end Sequences of Library MSMg01

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1469)

AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.

TITLE Direct Substitution

JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC),

1-7-22 Suenho-chou,Tsukuba-shi, Ibaraki, Japan

(E-mail:hattori@gsr.riken.jp, URL:http://hgp.gsc.riken.go.jp/)

Tel:81-45-503-9111, Fax:81-45-503-9170

Clones are derived from the mouse BAC library MSMg01. For BAC

library availability, please contact Kunihya Abe (abe@rtc.riken.jp).

Tsukuba Institute, Bio Resource Center,

The Institute of Physical and Chemical Research (RIKEN) 3-1-1

Koyada, Tsukuba, 305-0074 Japan

phone: 81-298-36-9189, fax: 81-298-36-9199

e-mail: abe@rtc.riken.jp

PRIMERS

Sequencing : T7

LIBRARY

Vector : PBACe3.6

R.Site 1 : EcoRI

R.Site 2 : SmaI

Location/Qualifiers

1..1469

/organism="Mus musculus molossinus"

/mol\_type="genomic DNA"

/sub\_species="molossinus"

/db\_xref="taxon:57486"

/clone="MSMg01-168A20.T7"

/sex="male"

/tissue\_type="mixture of kidney and spleen"

/clone\_lib="MSMg01 Mouse Male BAC Library"

## ORIGIN

Query Match 13.7%; Score 133; DB 9; Length 1469;

Best Local Similarity 47.8%; Pred. No. 1.1e-12;

Matches 408; Conservative 0; Mismatches 442; Indels 4; Gaps 1;

```
QY 9 CTTCCCGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 68
    |||||
DB 1384 CTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1325
QY 69 GCTGCGGAGGCGGCGGACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 128
    |||||
DB 1324 CMC CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1265
QY 129 GAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 188
    |||||
DB 1264 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1205
QY 189 GAGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 248
    |||||
DB 1204 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1145
QY 249 CGTGACCGGCTCATGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 308
    |||||
DB 1144 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1089
QY 309 CATGCGCGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 368
    |||||
DB 1088 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1029
QY 369 GTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 428
    |||||
DB 1028 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 969
    |||||
QY 429 GCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 488
    |||||
```







SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT  
FEATURES  
source

Pristionchus pacificus  
Pristionchus pacificus  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.  
1 (bases 1 to 1406)  
Srinivasan,J., Sins,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,  
Bunfeler,J., van der Meulen,W. and Sommer,R.U.  
An integrated physical and genetic map of the nematode Pristionchus  
pacificus  
Mol. Genet. Genomics 269 (5), 715-722 (2003)  
22835951  
12884007  
Contact: Sommer RU  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498  
Email: ralf.sommer@tuebingen.mpg.de  
Class: BAC ends.  
Location/Qualifiers  
1. .1406  
/organism="Pristionchus pacificus"  
/mol\_type="genomic DNA"  
/strain="California"  
/db\_xref="taxon:54126"  
/clone\_lib="Ppa EORI BAC library"  
/note="The library was generated by a partial digest of  
the genomic DNA with EcoRI and cloning into the BAC  
vector."

ORIGIN

Query Match 13.3%; Score 128.6; DB 9; Length 1406;  
Best Local Similarity 47.4%; Pred. No. 6e-12;  
Matches 427; Conservative 0; Mismatches 471; Indels 3; Gaps 2;

49 GCGCGATGTTCCGCGGCTGCTGCGGAGCGGCGACCGCTGCTGTCACCTC 108  
|||||  
421 GCGCGCGCGGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
|||||  
109 GTACCGCGCGCGGAGCGCGCGCGCTGCTGCTGCGCGCGCGCGCGCGCGCGCG 168  
|||||  
481 GCG 540  
|||||  
169 GAACCTGCG 228  
|||||  
541 CGGCG 600  
|||||  
229 GCG 288  
|||||  
601 GCG 660  
|||||  
289 ACCCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 347  
|||||  
661 CG 720  
|||||  
348 GCACTGCG 407  
|||||  
721 CG 780  
|||||  
408 CGCGGTGTCACGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 467  
|||||  
781 CG 840  
|||||  
468 CG 527  
|||||  
841 CG 900  
|||||  
528 GCGCGTGAAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 587  
|||||  
901 CG 960  
|||||  
588 CGTCCG 647  
|||||

Db 961 NCGCCG 1020  
|||  
Qy 648 CCGTGTGCTGCG 707  
|||  
Db 1021 GCG 1080  
|||  
Qy 708 GCGCGTCCG 767  
|||  
Db 1081 GCG 1138  
|||  
Qy 768 CGACCGGACCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 827  
|||  
Db 1139 CG 1198  
|||  
Qy 828 CGACCG 887  
|||  
Db 1199 CG 1258  
|||  
Qy 888 GCTCCG 947  
|||  
Db 1259 CG 1318  
|||  
Qy 948 C 948  
|||  
Db 1319 C 1319

RESULT 5  
AG381986/c 2243 bp DNA linear GSS 03-JUN-2004  
LOCUS  
DEFINITION  
Mus musculus molossinus DNA, clone:MSMg01-192D21.TU, genomic survey  
sequence.  
ACCESSION  
AG381986  
VERSION  
AG381986.1 GI:47993191  
KEYWORDS  
GSS.  
SOURCE  
Mus musculus molossinus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.  
BAC end Sequences of Library MSMg01  
Unpublished  
JOURNAL  
TITLE  
AUTHORS  
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.  
2 (bases 1 to 2243)  
DIRECT SUBMISSION  
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
Clones are derived from the mouse BAC library MSMg01. For BAC  
library availability, please contact Kumiya Abe (abe@tc.riken.jp).  
Tsukuba Institute, Bio Resource Center,  
The Institute of Physical and Chemical Research (RIKEN) 3-1-1  
Koydai, Tsukuba, 305-0074 Japan  
phone: 81-298-36-9189, fax: 81-298-36-9199  
e-mail: abe@tc.riken.jp  
PRIMERS  
Sequencing : TU  
LIBRARY  
Vector : pBAC3.6  
R.Site 1 : EcoRI  
R.Site 2 : EcoRI.  
FEATURES  
source  
1. .2243  
Location/Qualifiers  
/organism="Mus musculus molossinus"  
/mol\_type="genomic DNA"  
/sub\_species="molossinus"  
/db\_xref="taxon:57486"  
/clone="MSMg01-192D21.TU"  
/sex="male"







QY 419 CCAGGAGAGGGGCGGGGGCGTCAACGAGCCCTGCTGCGAGGCGCGCGCCGACAGC 478

Db 662 GGGGGCGGCGCGCGCGCGCGCGGGGCGGGGCGGGCCCGCGCGAGGGGGGGGGGCGCG 603

QY 479 CCGTACGGGCTCAACGCGCGAGGAGGACGACCGGACCGACGAGCGGACCCAGAGCGCTGACGC 538

Db 602 CGCGCGCGGCC 543

QY 539 ACGCGGTGTCCTCTTCCTTCCTTCGGGGCTCGCCCTTGCGCGGCTCGGGCTCGACGTTCGGGCCC 598

Db 542 GCC 483

QY 599 TGGCGGCGAGCGGACCGCGCGCGCCCAACAAGTGTCTGCGCGCTCTCGGCGCGGTGGCTCG 658

Db 442 CCGCC 423

QY 659 GCGGAGCGCCCGAGGTGTACGGGAGCATTCAGCGGTCCACCCTCGGGCGGGCGGTCCGCGC 718

Db 442 CCCCCCGCGCGCGCGCGGGGGGCGCGCGCGCTGCGCGCGGCGCGCGCGCGCGCGCGCC 363

QY 719 GCGGGGCGCTCGCGGAGCGCTTGCGCGCTCTTCCTTCGCGCGCGCTGCTGCGGACGACCCGGAC 778

Db 362 CGCGCGCGCGCGCGGGAGCGCGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 303

QY 779 GTTCCGAGCGCGCCCGGGGCGCGCGACGCGCGCGCGCGCGCATTCGCGGGGAGATTGCGACGCGCGCG 838

Db 302 GCGCGCGCGGCC 243

QY 839 GGAACCTTGAACGGCGCTCTTCGGGGAATTCCGCGGCTCATGGGACCGGAGCTCGC 893

Db 242 CNGCC 188

RESULT 7	
LOCUS	BM450237
DEFINITION	BM450237 1350 bp mRNA linear EST 05-FEB-2002
ACCESSION	AGENCOURT_6393396 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5528315
VERSION	5', mRNA sequence.
KEYWORDS	BM450237 BM450237.1 GI:18499277
SOURCE	EST.
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
JOURNAL	1 (bases 1 to 1350)
COMMENT	NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov Tissue Procurement: ATCC/DCTD/DTF cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLM12204 row: p column: 12 High quality sequence stop: 370.
FEATURES	Location/Qualifiers
source	1..1350
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone_image="IMAGE:5528315"
	/tissue_type="melanotic melanoma"
	/lab_host="DH10B (phage-resistant)"
	/clone_lib="NIH MGC 72"
	/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Library constructed by Life

ORIGIN	Technologies."
Query Match	13.1%; Score 126.8; DB 4; Length 1350;
Best Local Similarity	43.8%; Pred. No. 1.2e-11;
Matches 399; Conservative	0; Mismatches 507; Indels 5; Gaps 2;
Qy	42 GGGCGTGGGCGGCGCATGTTTCGCGGGGCTGCTGCGGAGAGCGGGACGCCGACGCTGTGCT 101
Db	286 GCGTGTGTGGTGCATCTCTCTTTCACGCGCTCTCCCTCCAGATGGAGATGTCGCG 345
Qy	102 CGACCTCGTACCGCGCGCGGGACCGCGGACGCGCTGCTGTGGGCGACGCTACCGCGCC 161
Db	346 TGGACCGGAGCGCGGGCGGNN 405
Qy	162 GGGGCGCGAACTCGCGCGCGCGCTCCGCGGACGCGGACCTGCTCTGCTGCGCGGTACAGA 221
Db	406 GCGCGCGCGCGCGCGGGGGCG 465
Qy	222 GGAAGTGGCGCTCAAGGCTGTGGGCGCGCGGTACCGCGCTATGTGGCGCGGGCGGCTGCT 281
Db	466 GNCG 525
Qy	282 CGCGGACACCTGTCCGTCCGCGGACGGGCGATGGCGCGGAGTGTGCGCGCGCCACGCGCGCG 341
Db	526 GNGCGANCGNNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 585
Qy	342 CGTCCAGACGCTGGGCGCTCAACCCGATGTTTCGCGCGCGCGCGCGCGCGCGCGCGCGAC 401
Db	586 CCG 645
Qy	402 CGTGGCGCGCGCTGTGTACACAGGAGCGGGCGCGGCGTACCGGCGCTGCTGTGCGCTGTG 461
Db	646 GGGNCG 704
Qy	462 GGGGCGGGGCGGGGCGCGCGCGCTCACGGCGGAGAGACGACCGGAGCGAGCGGGCG 521
Db	705 GGGCGCGCGCGGGCGCGCGCGGNGGNGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 764
Qy	522 CACCCAGGCGCTGTACGACGCGCGTGTCTCTCTTGGGAGTGTGCGCGCGCGCGCTGTG 581
Db	765 CGGCG 824
Qy	582 CGTGCAGCTGCG---GACCTGTGGCGGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 637
Db	825 CG 884
Qy	638 CCGTCTGTGGCGCGGT 697
Db	885 CCG 944
Qy	698 ACCCGCGGCGCGCGCTGCG 757
Db	945 GGGCG 1004
Qy	758 TGTGTGCGGACGACCGCGGACCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 817
Db	1005 CCG 1064
Qy	818 CCGGGGGAATGTGAGCGGCGCGCGGAACTTCGACGCGGCTTTCGCGGGAATCTCGCGGCTCA 877
Db	1065 CCG 1124
Qy	878 TGGACCGGAGCTGCGCGCGCGCGCAGAGACGACGCGGAGGCTGTTCGCGACCTTCACG 937
Db	1125 CG 1184
Qy	938 GCACCGGACGAC 948
Db	1185 NGCGNCGCGC 1195



LOCUS	AC448243	1780 bp	DNA	linear	GSS 03-JUN-2004
DEFINITION	AG448243	Mus musculus molossinus DNA, clone:MSWg01-332C01.T0, genomic survey sequence.			
ACCESSION	AG448243				
VERSION	AG448243.1	GI:48091306			
KEYWORDS	GSS.				
SOURCE	Mus musculus molossinus				
ORGANISM	Mus musculus molossinus				
REFERENCE	Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.				
AUTHORS	BAC end Sequences of Library MSWg01				
TITLE	Unpublished				
JOURNAL	2 (bases 1 to 1780)				
REFERENCE	Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.				
AUTHORS	Direct Submission				
TITLE	Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Shuhiro-Chou, Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:shattori@gsc.riken.jp, URL:http://hgp.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)				
JOURNAL	Clones are derived from the mouse BAC library MSWg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).				
COMMENT	Yoshida Institute, Bio Resource Center, The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan phone: 81-298-36-9189, fax: 81-298-36-9199 e-mail: abe@rtc.riken.jp				
PRIMERS					
Sequencing	: TJ				
LIBRARY	Vector	: pBACe3.6			
Library	R.Site 1	: EcoRI			
R.Site 2	: EcoRI.				
FEATURES					
source	Location/Qualifiers				
	1..1780				
	/organism="Mus musculus molossinus"				
	/mol_type="genomic DNA"				
	/sub_species="molossinus"				
	/db_xref="taxon:57486"				
	/clone="MSWg01-332C01.TJ"				
	/sex="male"				
	/tissue_type="mixture of kidney and spleen"				
	/clone_lib="MSWg01 Mouse Male BAC Library"				
ORIGIN					
Query Match	13.1%;	Score 126.6;	DB 9;	Length 1780;	
Best Local Similarity	48.7%;	Pred. No. 1.3e-11;			
Matches 461;	Conservative 0;	Mismatches 466;	Indels 18;	Gaps 5;	
OY	3	GAGCGGCTTCCCCCGACGCTGTCGTGGCGCGCAGCGGGCGGTGGCGGCATGTTGC	62		
Db	579	GGGTCGCGCCCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	638		
OY	63	CGGGCTGCTGCGGAGAGCGG-----GAGCGCGCAGCGTCTGTCGACCTCGAACCGCGC	118		
Db	639	CGGCG	698		
OY	119	CGGAGCGGCGCGGACGCTGCTGTCGTGGCGCAGTCGACCGCGCGCGGAGCGCGAATTGCGG	178		
Db	699	GCGGCGCGCGCGCGCGCTTTTTCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	758		
OY	179	CGGCGCTCCGGGACCGCGACCTGTCGTCTGCTGTCGCGCTACAGAGACGTGGCCCTCAAG	238		
Db	759	CG	818		
OY	239	CGGTGGCGCGCGTGAACCGGCTCATGCGGCGCGCGCGCGCGCTGCTCGCGACACCTGTGCG	298		
Db	819	CG	874		
OY	299	TCGGGACGGGCGATGGCGCGCGAGCTCGCGGCGCGCGCGCGCGCGCGCGCGCTTCAGACGTGGCG	358		

[illegible]











[illegible]

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
CL090560/c	1821 bp DNA	CL090560	1	linear	GSS 05-JAN-2004	
	ISB1-17N15.T7.1 ISB1 <i>Xenopus tropicalis</i> genomic clone ISB1-17N15, genomic survey sequence.	CL090560				
		CL090560.1	GI:40584195			
		GSS.				
	<i>Xenopus tropicalis</i> (western clawed frog)					
	<i>Xenopus tropicalis</i>					

REFERENCE	1 (bases 1 to 1821)
AUTHORS	Kremliukh, C., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E. and Wilson, R.
TITLE	A physical map of the xenopus tropicalis genome
JOURNAL	Unpublished (2003)
COMMENT	Contact: Richard K Wilson

Washington University School of Medicine  
Email: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
Insert Length: 75000 Std Error: 0.00  
Seq primer: T7 TAATACGACTCACTATAGGG  
Class: BAC ends  
High quality sequence start: 591  
High quality sequence stop: 665.

FEATURES	source	Location/Qualifiers
1..1821		
/organism="Xenopus tropicalis"		
/mol_type="genomic DNA"		
/db_xref="taxon:8364"		
/clone="ISB1-17N15"		
/clone_1fb="ISB1"		
/note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC Library Segment 1"		

	ORIGIN	
Query Match	12.9%	Score 125.2; DB 9; Length 1821;
Beet Local Similarity	47.8%;	Pred. No. 2.2e-11;
Matches 436;	0; Mismatches 463;	Indels 13; Gaps 2;
Oy	17	GAGCGTCGTCGTCCGACGAGCGGGGGTGAGCGGCAGATGTCGCCCGGGCTGCTGCGGG 76

[illegible]

LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE
CH216-16011	1218 bp	DNA	linear	GSS 31-DEC-2003
CH216-160116	CH216 <i>Xenopus tropicalis</i> genomic clone			
CH216-160116	genomic survey sequence.			
CH216-16011				
CH216-16011.1	GI:40536924			
GSS.				
<i>Xenopus tropicalis</i>	(western clawed frog)			



ORGANISM	REFERENCE AUTHORS	TITLE JOURNAL COMMENT
Xenopus tropicalis	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.	
1 (bases 1 to 1218)	Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E., and Wilson, R.	A physical map of the xenopus tropicalis genome Unpublished (2003)
CONTACT: Richard K Wilson		
Genome Sequencing Center		
Washington University School of Medicine		
Email: submissions@wustl.edu		
Insert Length: 17500		
Std Error: 0.00		
Seq primer: RM4 ctcaaggacacggtcagc		
Class: BAC ends		
High quality sequence start: 79		
High quality sequence stop: 120.		
Location/Qualifiers		
1. 1218		
/organism="Xenopus tropicalis"		
/mol_type="genomic DNA"		
/strain="Nigerian frog"		
/db_xref="taxon:8164"		
/clone="CH216-160116"		
/sex="male"		
/cell_line="Stock 248 F7A2, inbred N7"		
/clone_lib="CH216"		
/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis BAC library"		
ORIGIN		
Query Match	12.9%; Score 124.6; DB 9; Length 1218;	
Best Local Similarity	43.6%; Pred. No. 2.8e-11;	
Matches 405; Conservative	0; Mismatches 519; Indels 4; Gaps 1;	
18 CAGCGTCGTCGTCGCGCGAGCGGGGCGGCGGATGTTCCGCGGCGTCTGCGGGA	77	
222 CGCGCGCGGAGNNNGGCGCNCNCGCGGGGGCGCGCGGCGCNCNGGCGCGCGGAGGCG	291	
78 GCGCGGAGCGCGGACGCTCGTCGTGCGACTGTACCGCGCGCGCGCGCGCGGACGCTG	137	
292 CGCGCGCGCGCGCGCGCGGGGAGNNNGCGCCNNNGCGCGCGCGCGCNCNCCGCGCNCNGG	351	
138 CCGTGTGGGAGCGTCAACGCGCGCGGGGCGCGAACTCGCGCGCGCTTCCGGAGCGCGGA	197	
352 CCG	411	
198 CCGTGTCTGCTCGCGCTGACACGAGGAGCGTGGCGCTCAAGGCGCGTGGCGCGCGTGA	257	
412 CCG	471	
258 GCGTATGCG	317	
472 NNNNNCCCGCGCGCGCGGGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	531	
318 GAGGTCG	377	
532 GCG	591	
378 CG	437	
592 CCG	651	
438 CAGCG	497	
652 CCG	711	
498 GAGGACGAGCG	557	
712 GCG	771	
558 CGGCG	617	

ORIGIN



```
Query Match      12.8%; Score 124.6; DB 9; Length 1267;
Best Local Similarity 47.1%; Pred. No. 2.8e-11;
Matches 399; Conservative 0; Mismatches 448; Indels 1; Gaps 1;

QY 79 GCGGGACGCGGACGCTGCTGCTGACCTGACCGCGCGGGAGACGCGCGCTGC 138
DB 1191 GCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1132
QY 139 CTGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 198
DB 1131 GCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1072
QY 199 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 258
DB 1071 GCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1012
QY 259 CTGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 318
DB 1011 TNGCGCGCGG-GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 953
QY 319 GAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 378
DB 952 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 893
QY 379 GCGCGCGCGCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 438
DB 892 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 833
QY 439 AGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 498
DB 832 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 773
QY 499 GAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 558
DB 772 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 713
QY 559 GGGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 618
DB 712 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 653
QY 619 CCGCAGCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 678
DB 652 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 593
QY 679 GGGAGCATCCAGCGCGCTGCAACCGCGCGCGCGCGCGCGCGCGCGCG 728
DB 592 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 533
QY 739 CTGCGCTCTTTCGCGCGCGCTGCTGCGCGAGCAGCCGCGCGCGCGCG 798
DB 532 MCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 473
QY 799 GCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 858
DB 472 GCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 413
QY 859 GGGAGACTCCGCGCGCTGCTGAGCAGGAGCTGCGCGCGCGCGCGCG 918
DB 412 CCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 353
QY 919 CTGTTCCG 926
DB 352 CCGGCGCG 345

RESULT 15
AG363333/c      2332 bp      DNA      linear      GSS 03-JUN-2004
LOCUS           Mus musculus molossinus DNA, clone:MSMg01-165A22.TJ, genomic survey
DEFINITION
ACCESSION       AG363333
VERSION         AG363333.1      GI:47974538
KEYWORDS        GSS.
```

```
SOURCE
ORGANISM        Mus musculus molossinus
MUS musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS        Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE          BAC end Sequences of Library MSMg01
JOURNAL        Unpublished
REFERENCE
AUTHORS        Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE          Direct Submission
JOURNAL        Submitted (17-NOV-2003) Maehara Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suenho-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT        Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kunita Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyada, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : PBAC3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES
source
1..2332
Location/Qualifiers
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-165A22.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN
Query Match      12.8%; Score 124.4; DB 9; Length 2332;
Best Local Similarity 48.8%; Pred. No. 2.8e-11;
Matches 421; Conservative 0; Mismatches 436; Indels 5; Gaps 3;

QY 83 GAGCGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 142
DB 2163 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2104
QY 143 TGGGCGACGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 202
DB 2103 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2044
QY 203 TCTGCTGCGCGTCAACGAGAGCTGAGCGCG--TGGCGCGGTGACCGCGC 259
DB 2043 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1984
QY 260 TCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 319
DB 1993 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1924
QY 320 AGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 379
DB 1923 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1864
QY 380 CCGCGCGCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 439
DB 1863 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1804
QY 440 CCGCGCTGCTGCGCGCTGCTGCGAGCGCGCGCGCGCGCGCGCGCGCG 499
DB 1803 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1744
```



```
QY 500 AGACAGACCGGACGACCGGCGCCACCCAGGCCCTGACGACGCGGTGCTCCTTCCTTCG 559
Db 1743 GCGCGGATCGCGCCCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1684
QY 560 GAGTCGCGCTTCGCGCGCTTCGCGCTTCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCG 619
Db 1683 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1625
QY 620 CCGACCGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 679
Db 1624 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1565
QY 680 GAGACATTCAGCGGATTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 739
Db 1564 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1505
QY 740 TCGCGCTCCTTCGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 799
Db 1504 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1446
QY 800 CCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 859
Db 1445 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1386
QY 860 GAGAACTTCGCGCGCGCTCATGAGACCGGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 919
Db 1385 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1326
QY 920 TGTTCGCGACCTTCACCGCGAC 941
Db 1325 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1304
```

Search completed: October 5, 2005, 09:01:18  
Job time : 3671.67 secs



This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 05:11:11 ; Search time 4767.69 Seconds  
(without alignments)  
3272.568 Million cell updates/sec

Title: US-10-089-514-6  
Perfect score: 1635  
Sequence: 1 MSGFPRSVVVGSGAVGCMF.....HCOELFRTLHRTDDEGKDR 322

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues  
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame\_p2n.model -DEV=x1h  
-Q=/cg22\_1/USPTO.epool\_h/US10089514/runat\_04102005\_105744\_8033/app\_query.fasta\_1.1621  
-DB=GenBdb1 -QFMT=faetap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LIST=45  
-UNITS-bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=psio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USFR=US10089514\_@CGN\_1\_1\_8782\_@runat\_04102005\_105744\_8033 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -PGAPOP=6  
-PGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenBdb1:\*  
1: gb\_da:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	1635	100.0	969	BD178315	BD178315 Transform
2	1635	100.0	969	BD093916	BD093916 Transform
3	1635	100.0	5251	AB116234	AB116234 Streptomy
4	1588	97.1	14159	AF262220	AF262220 Streptomy

# SUMMARIES

5	661.5	40.5	888	6	A48324	A48324 Sequence 2	
6	661.5	40.5	888	6	AR198354	AR198354 Sequence	
C	7	661.5	40.5	2888	A48323	A48323 Sequence 1	
C	8	661.5	40.5	2888	AR198353	AR198353 Sequence	
C	9	661.5	40.5	4740	SP060417	SP060417 Streptomyces	
C	10	382	23.4	349944	1 BX571871	BX571871 Photocyc	
11	382	23.4	349980	6	AX770909	AX770909 Sequence	
12	323	19.8	14192	1	AE010371	AE010371 Methanopy	
13	272	16.6	12180	1	AE011181	AE011181 Methanosa	
14	256	15.7	9888	1	AE013358	AE013358 Methanosa	
15	239	14.6	268122	1	BX957223	BX957223 Methanoco	
C	16	237.5	14.5	21387	1	AE008821	AE008821 Salmonell
C	17	237.5	14.5	274050	1	AL627276	AL627276 Salmonell
C	18	237.5	14.5	300029	1	AE016842	AE016842 Salmonell
C	19	237	14.5	10327	1	AE015281	AE015281 Shigella
C	20	237	14.5	12071	1	AE005490	AE005490 Escherich
C	21	237	14.5	270365	1	AP002562	AP002562 Escherich
C	22	237	14.5	290380	1	AE016987	AE016987 Shigella
23	235	14.4	9739	1	U67509	U67509 Methanocald	
C	24	235	14.4	110000	6	AR271569_09	Continuation (10 o
C	25	234	14.3	300099	1	AE016764	AE016764 Escherich
C	26	233	14.3	1122	6	AX457133	AX457133 Sequence
27	233	14.3	1238	6	AX343931	AX343931 Sequence	
28	233	14.3	1238	6	AX417716	AX417716 Sequence	
C	29	233	14.3	4509	1	ECOPHEAB	M10431 E.coli phe
C	30	233	14.3	19338	1	D90888	D90888 E.coli geno
C	31	233	14.3	110000	1	U00096_27	Continuation (28 o
C	32	230	14.1	11696	1	AE000923	AE000923 Methanoba
C	33	230	14.1	110000	1	BX950851_37	Continuation (38 o
34	228	13.9	11501	1	AE015580	AE015580 Shewanell	
35	227	13.9	1131	6	AR387843	AR387843 Sequence	
36	225.5	13.8	1012	1	ERWTYRA	M74135 Erwina her	
37	225.5	13.8	1152	6	AX457135	AX457135 Sequence	
C	38	225.5	13.8	3434	1	EHPEAGN	X60420 Erwina herb
C	39	225.5	13.8	3434	1	ERWPTA	M74133 Erwina her
C	40	224	13.7	3957	6	CO872837	CO872837 Sequence
C	41	224	13.7	349980	6	CO873160	CO873160 Sequence
42	219.5	13.4	11069	1	AE013693	AE013693 Yersinia	
43	219.5	13.4	110000	1	BX936398_10	Continuation (11 o	
C	44	219.5	13.4	220050	1	AJ414156	AJ414156 Yersinia
45	219.5	13.4	290803	1	AE017128	AE017128 Yersinia	

## ALIGNMENTS

RESULT 1	BD178315	969 bp	DNA	linear	PAT 16-APR-2003
LOCUS	BD178315	Transformant producing pRI022 substance and process for producing the same and novel biosynthetic gene.			
DEFINITION	BD178315	the same and novel biosynthetic gene.			
ACCESSION	BD178315.1	GI:30015580			
VERSION	BD178315	WO 02077244-A/3.			
KEYWORDS	Streptomyces venezuelae				
SOURCE	Streptomyces venezuelae				
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomyces				
REFERENCE	1 (bases 1 to 969)				
AUTHORS	Yanai, K., Sumida, N., Watanabe, M., Moriya, T. and Murakami, T.				
TITLE	Transformant producing pRI022 substance and process for producing the same and novel biosynthetic gene				
JOURNAL	Patent: WO 02077244-A 3 03-OCT-2002;				
COMMENT	MEIJI SEIKA KAISHA LTD, KOJI YANAI, NAOMI SUMIDA, MANABU WATANABE, TATSUKI MORIYA, TAKESHI MURAKAMI				
OS	Streptomyces venezuelae				
PN	WO 02077244-A/3				
PD	03-OCT-2002				
PF	22-MAR-2002	WO 2002JP002782			
PR	22-MAR-2001	JP 01P 082227			
PI	KOJI YANAI, NAOMI SUMIDA, MANABU WATANABE, TATSUKI MORIYA, TAKESHI MURAKAMI				
PC	C12N15/61, C12N1/15, C12P17/14, C12N1/15, C12R1.645, C12P17/14,				



PC C12R1:645)  
CC Transformant producing PF1022 substance and process for CC  
CC producing the same  
CC and novel biosynthetic gene  
FH Key Location/Qualifiers  
FT CDS (1)..(966).  
FEATURES  
source 1..969  
/organism="Streptomyces venezuelae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:54571"  
ORIGIN  
Alignment Scores:  
Pred. No.: 2e-62 Length: 969  
Score: 1635.00 Matches: 322  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0  
US-10-089-514-6 (1-322) x BD178315 (1-969)  
QY 1 MetSerGlyPheProArgSerValValGlyGlySerGlyAlaValGlyGlyMetPhe 20  
DB 1 ATGAGCGGCTTCCCGCGACGTCGTCTCGCGCGGACGCGGGCGGTGGCGGCGATGTTT 60  
QY 21 AlaGlyLeuLeuArgGlyValGlySerArgThrLeuValAlaAspLeuValProProPro 40  
DB 61 GCGGGGCTGTGGGGGAGCGGGGACGCGGACCTGCTGCACCTCGTACCGCGCGCCG 120  
QY 41 GlyArgProAspAlaCysLeuValGlyAspValThrAlaProGlyProGlyLeuAlaAla 60  
DB 121 GAGCGGCGGACCGCTCGTGGGGGAGTCACGCGCGCGCGGCGCGGACCTCGCGGC 180  
QY 61 AlaLeuArgAspAlaAspLeuValLeuLeuAlaValHisGlyAspValAlaLeuValAla 80  
DB 181 GCCCTCGGACCGCGGACCTGCTGCTGCTGCGGTACAGAGACGTCGCGCTCAAGGCC 240  
QY 81 ValAlaProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerVal 100  
DB 241 GTGGCGGCTGTGACCGGCTCATGCGCGCGGGGCGGTGCTGCTCGACACCTGTGCTCGTC 300  
QY 101 ArgThrGlyMetAlaAlaGlyLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeu 120  
DB 301 CGAGCGGACATGGCGCGGAGCTCGGGGCCACGCGCGCGCTCCAGACGTCGGGCTTC 360  
QY 121 AsnProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValAlaThr 140  
DB 361 AACCCATGTTCCGCGCGCGCGGACATGACCGCGCGGACCCGTGGCGCGCTGCTCACCC 420  
QY 141 ArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGlnGlyGlyArgProPro 160  
DB 421 AGGAGAGGGGCGGGCGGTACGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
QY 161 ValArgLeuThrAlaGlnGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHis 180  
DB 481 GTACGGCTACCGCGGAGAGGACGACGCGGACGCGGCGGCGGCGGCTGAGCGCAC 540  
QY 181 AlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValAlaGlyAlaLeu 200  
DB 541 GCCGTCTCTCTCTCTCGGGGCTCGGCTCGCGCGCTCGGCGGCTCGGCGGCGGCTG 600  
QY 201 AlaAlaThrAlaProProProHisGlnValAlaLeuAlaLeuLeuAlaArgValLeuGly 220  
DB 601 GCGGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
QY 221 GlySerProGlyValValArgAlaPheGlnArgSerAspProArgAlaAlaSerAlaArg 240  
DB 661 GCGAGCGCGGAGGTGACGGGACATCCAGCGGTCCACCGCGGCGGCGGTCCGCGCGC 720  
QY 241 ArgAlaLeuAlaGlyAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArg 260

DB 721 CGGGCGCTCGCGGAGCGCTTCGCGCGGCTGTCGCGACGACCGCGGACCGT 780  
QY 261 AlaAspAlaProGlyValArgAlaAspAlaProGlyHisProGlyGlyValCysAspGlyAlaGly 280  
DB 781 GCGGAGCGCGCGGGGCGCGGACGCGCGCGGACATCCCGGGGAGATGCGAGCGCGCGG 840  
QY 281 AsnLeuAspGlyValPheGlyGlyLeuLeuArgArgLeuMetGlyProGlyLeuAlaAlaGly 300  
DB 841 AACCTGACGCGGCTTTCGGGGGAACTCGCGCGGCTCATGGAACCGGAGCTCCGCGCGGCG 900  
QY 301 GlnAspHisCysGlnGlyLeuLeuPheArgThrLeuHisArgThrAspAspGlnGlyValGly 320  
DB 901 CAGGACCACTGCACAGAGCTGTTCGCGACCTTCACCGGACGACGACGAGCGAGAG 960  
QY 321 AspArg 322  
DB 961 GACCGA 966  
RESULT 2  
BD093916 969 bp DNA linear PAT 27-AUG-2002  
LOCUS  
DEFINITION Transformants that produce secondary metabolites modified by a functional group(s) and novel biosynthesis genes.  
ACCESSION BD093916.1 GI:22639504  
VERSION BD093916.1  
KEYWORDS WO 0123542-A/3.  
SOURCE  
ORGANISM Streptomyces venezuelae  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomyces; Streptomyces; Streptomyces; Streptomyces.  
REFERENCE 1 (bases 1 to 969)  
AUTHORS Yanai,K., Okakura,K., Yasuda,S., Watanabe,M., Miyamoto,K., Mido,N. and Murakami,T.  
TITLE Transformants that produce secondary metabolites modified by a functional group(s) and novel biosynthesis genes  
JOURNAL Patent: WO 0123542-A 3 05-Apr-2001;  
MEIJI SEIKA KAISHA LTD, KOJI YANAI, KAORU OKAKURA, SHOHEI YASUDA, MANABU WATANABE, KOICHI MIYAMOTO,  
OS Streptomyces venezuelae  
COMMENT PN WO 0123542-A/3  
PD 05-APR-2001  
PF 29-SEP-2000 WO 2000JP006783  
PI 29-SEP-1999 JP 99P 276314  
PI KOJI YANAI, KAORU OKAKURA, SHOHEI YASUDA, MANABU WATANABE, KOICHI MIYAMOTO,  
PI NAKOKI MIDO, TAKESHI MURAKAMI  
PC C12N15/09, C12N5/10, C12P21/02, C07K11/00// (C12P21/02, C12R1:645)  
CC  
FH Key Location/Qualifiers  
FT CDS (1)..(966).  
FEATURES  
source 1..969  
/organism="Streptomyces venezuelae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:54571"  
ORIGIN  
Alignment Scores:  
Pred. No.: 2e-62 Length: 969  
Score: 1635.00 Matches: 322  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0  
US-10-089-514-6 (1-322) x BD093916 (1-969)  
QY 1 MetSerGlyPheProArgSerValValGlyGlySerGlyAlaValGlyGlyMetPhe 20  
DB 1 ATGAGCGGCTTCCCGCGACGTCGTCTCGCGCGGACGCGGGCGGTGGCGGCGATGTTT 60  
QY 21 AlaGlyLeuLeuArgGlyValGlySerArgThrLeuValAlaAspLeuValProProPro 40







```
/db_xref="GI:47846868"
/translation="MTEQNDLQRLRAELDALDGLTLDTVRRRIDLVRIARYKSRHGV
PMQGRVSLVKDRAARAYADHGLDESLVLYDLITTEMCREDVLMSRESLTADR
R"
4065..5243
/gene="ORF1V"
4065..5243
/gene="ORF1V"
/codon_start=1
/evidence=not experimental
/translation_table=11
/product="hypothetical protein"
/protein_id="BAD21143.1"
/db_xref="GI:47846869"
/translation="MAHTTTLAALGTGRTAAQCPYPMPDPAVDTERELAGLPVLTPEP
EVDLNMGMALVARGAALIVGGDCAERFHEAVPDIVRRKVDNLQGLAIMRAGSLP
AVALGRLAGOYGRSPSPPTESPESGAKRKKPKYCGAVNPEFEPARLPPELLTA
YTCSSQIVDEIRSRMSGRPLERYTSHLELLPYERPLVREAGAGTSGSAHFHMG
EKTRREDGAHVALAQAVHNVPVGLPTVSBDAVALSRVLENGVGRLLTIVRGA
KEVDLPLPVRAVARHGAPEVWLCDPMHNGKLGHKTRLTEPMRAETAAVRLTIR
EHQMPAGLHLELTPDPTVECVSELDPRPFTDTRSTCDPRLNPEQASADWTFHSLT
"
```

## ORIGIN

## Alignment Scores:

```
Pred. No.: 9.25e-62 Length: 5251
Score: 1635.00 Matches: 322
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
```

US-10-089-514-6 (1-322) x AB116234 (1-5251)

```
QY 1 MetSerGlyPheProArgSerValValValGlyGlySerGlyValValGlyGlyMetPhe 20
DB 2708 ATGAGCGGCTTCCCGCCGACGTCGTCTGCGGCGCGCGCGGCGGCGGCGGCGGCGGCTTCC 2767
QY 21 AlaGlyLeuLeuArgGluAlaGlySerArgThrLeuValValAspLeuValProProPro 40
DB 2768 GCCGGGCTGCTCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2827
QY 41 GlyArgProAspAlaCysLeuValGlyAspValThrAlaProGlyProGluLeuValAla 60
DB 2828 GGCACGGCGGACCGCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2887
QY 61 AlaLeuArgAspAlaAspLeuValLeuLeuValAlaHisGluAspValAlaLeuValAla 80
DB 2888 GCCCTCCGGGACGGGACGTCGTCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2947
QY 81 ValAlaProValThrArgLeuMetArgProGlyValAlaLeuLeuValAspThrLeuSerVal 100
DB 2948 GTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3007
QY 101 ArgThrGlyMetAlaAlaGluLeuValAlaHisAlaProGlyValGlnHisValGlyLeu 120
DB 3008 CGACGCGGCGATGCGCGGCGGAGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3067
QY 121 AspProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValAlaThr 140
DB 3068 AACCCGATGTTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3127
QY 141 ArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGluGlyGlyGlyArgPro 160
DB 3128 AGGGACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3187
QY 161 ValArgLeuThrAlaGluGluHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHis 180
DB 3188 GTACGCGCTCACGGGCGGAGACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3247
QY 181 AlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeu 200
DB 3248 GCCGTGCTCTCTCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3307
```

```
QY 201 AlaAlaThrAlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGly 220
DB 3308 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3367
QY 221 GlySerProGluValTyrGlyAspLleGlnArgSerAspProArgAlaAlaSerAlaArg 240
DB 3368 GGCAGCGCCCGAGGTGTACCGGGGACATCCAGCGGTCCAAACCCCGGCGGCGGCGGCGGCGG 3427
QY 241 ArgAlaLeuAlaGluValaLeuArgSerPheAlaAlaLeuValGlyValAspAspProAspArg 260
DB 3428 CGGCGCTCGCGAGGCGCTTGCGCTCTTGCGCGCGCTGTGCGGCGACGACGACCGGCGGCGG 3487
QY 261 AlaAspAlaProGlyValArgAlaAspAlaProGlyHisAspProGlyGlyCysAspGlyAlaGly 280
DB 3488 GCGGACGCGCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3547
QY 281 AsnLeuAspGlyValPheGlyGluLeuValArgLeuMetGlyProGluLeuAlaAlaGly 300
DB 3548 AACCTCGACGCGGCTCTTCGCGGGAACCTCCCGGCTCATGCGGACCGGAGCTTCGCGCGGCG 3607
QY 301 GlnAspHisCysGlnGluLeuPheArgThrLeuHisArgThrAspAspGluGlyGlyVal 320
DB 3608 CAGGACCACTGCGGAGGAGCTGTTCGACACCTTCACCGGACGACGACGACGACGACGACGAC 3667
QY 321 AspArg 322
DB 3668 GACCGA 3673
```

RESULT 4  
LOCUS AF262220 14159 bp DNA linear BCT 09-JAN-2002  
DEFINITION Streptomyces venezuelae chloramphenicol biosynthetic gene cluster,  
partial sequence.

ACCESSION AF262220  
VERSION AF262220.2 GI:14290414  
KEYWORDS

## SOURCE

Streptomyces venezuelae  
Streptomyces venezuelae  
Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales;

## REFERENCE

1 (bases 1 to 14159)  
He, U., Magarvey, N., Pirae, M. and Vining, L. C.  
The gene cluster for chloramphenicol biosynthesis in Streptomyces  
venezuelae ISP5230 includes novel shikimate pathway homologues and  
a monomodular non-ribosomal peptide synthetase gene  
Microbiology 147 (Pt 10), 2817-2829 (2001)

## JOURNAL

MEDLINE  
21461106  
PUBMED  
11577160

## REFERENCE

2 (bases 1 to 14159)  
He, U., Magarvey, N. A. and Vining, L. C.  
Direct Submission  
Submitted (01-MAY-2000) Biology, Dalhousie University, 1355 Oxford  
St, Halifax, NS B3H 4U1, Canada  
3 (bases 1 to 14159)  
He, U., Magarvey, N. A. and Vining, L. C.  
Direct Submission  
Submitted (05-JUN-2001) Biology, Dalhousie University, 1355 Oxford  
St, Halifax, NS B3H 4U1, Canada

## REFERENCE

AUTHORS  
TITLES  
JOURNAL

## REMARK

Sequence update by submitter  
On Jun 5, 2001 this sequence version replaced gi:10716943.  
FEATURES  
Location/Qualifiers  
1..14159

## source

/organism="Streptomyces venezuelae"  
/mol\_type="genomic DNA"  
/strain="ISP5230"

## CDS

/db\_xref="taxon:54571"  
/map="adjacent to pabAB"  
233..3511  
/function="essential for chloramphenicol biosynthesis"  
/codon\_start=1  
/translation\_table=11  
/product="CmlG"



## CDS

/protein\_id="AAC21974.1"  
/db\_xref="GI:10716944"  
/translation="MTAGSGPGRGTPLRPALAVRDIHSFIADLGMDCVNRCTAV  
GTYPRGVEDPAMREFVAVLGLTRALLADLVEAEEDVVLTPDRPGVGGPRLAD  
LVTAHNSRPLRIIDEAVRLA YWRHTDEYLVFVPSHL IYDLADVDLLQIRAHNLG  
DTPGRIGSRVRRHARTVEQYTLPAADPAEVLALBELVPGKGIPTISASVEWL  
PLEGVTPEDELSAVTGVLLPALGGGLVLOVYSRMEFAFGYVYIKLVVVRGP  
SDYTPYEAKTRHAEISLGRFTMSDLAPAPAFSGMRPVVUSDTTFMMPDRMW  
VPVRSARVPEDLKLADRSWPGPPLLMQYKRGFLSGAALADVDLGLKQIRGASGGTP  
"  
3498..6404  
/note="non-ribosomal peptide synthetase; contains  
reductase-like domains near C-terminus"  
/codon\_start=1  
/transl\_table=11  
/product="CmlH"  
/protein\_id="AAQ21975.2"  
/db\_xref="GI:14327897"  
/translation="MELPDLITLDLFRHRGRTRALRTAGRTWTEELDVRVTSALARID  
AECPRGRVIVAGEHTAEAVVWMLAARSHAVTPMNPQIPADRFEEPRVADAALIV  
CFERREALVRSERKAGRLIYAGDVGMPTDPAPAPADGTADPAPSRVAYSIGTSGSTD  
PKLVDSHGGLNLCSRLRLDITTPDQVLPASISFDASVEILGTLYAGATLVV  
VRDQASWLGVSRLAAGCDLMLSPSVARLDEARSRIKVEFCGALGGEYDK  
AARYSRVNAVGPTEATVCFSLAELTSTPSIGTPVDFPAPAYADPDSGDHATGSE  
LVIVGDVALGYAGSPAEENVEGTVDSPAVATGVVSLSDGELTYLRIDBIKR  
LGRHVNLAHVGSITSLRHGREVALVRDADTILVTADGATESLMARIQDLVPWE  
APDRVLVVDLPLTSGSKVRSALREILASPAAGHGCTGDEDAELRLVDTAVL  
GOELGEPSTLFDAGSSLAMIQIVKLSIDYGEBAVAAANADYDAPAFRLHNG  
EAVAPASAVDTPLRLRVERTERDLRAELPLLRDTRREPVGAADGRREVLTGASGF  
IGGHVLDRLAAGRPVLVSTGDDGVLTGATRGFGQADYARVRAISVAELERWD  
RRGPVVDVAVHCGGYVNHLLPLDSHLSGVNMTALVRAAALGASPAFLSAASAG  
ADRLPLSAATLVATGDPYRSKLISEYVNTLAVLCAVSHYRGLVYGRPEDRHL  
KDMFTRLLETARVGMFRLSGHVPICDGTLDLTLGRPDANPGDASRTPDSR  
SAVVNHTYALDELRLRHTGLTADVLA PAMFERVGRGVPAPVLLAAMALSGPM  
PSAHRVDHDLRLGLTPTPDTPAGDRPERTGTTSRPERRRPADTALTGTP"  
3576..3596  
/note="Region: adenylation domain A1"  
3708..3746  
/note="Region: adenylation domain A2"  
3939..3986  
/note="Region: adenylation domain A3"  
4089..4100  
/note="Region: adenylation domain A4"  
4350..4370  
/note="Region: adenylation domain A5"  
4494..4538  
/note="Region: adenylation domain A6"  
4593..4610  
/note="Region: adenylation domain A7"  
4647..4706  
/note="Region: adenylation domain A8"  
4833..4853  
/note="Region: adenylation domain A9"  
4893..4910  
/note="Region: adenylation domain A10"  
5043..5075  
/note="Region: thiolation domain"  
5343..5375  
/note="Region: NADH binding site"

## ORIGIN

Alignment Scores:  
Pred. No.:  
Score:  
Percent Similarity:  
Best Local Similarity:  
Query Match:  
DB:

2.38e-59  
1588.00  
96.65%  
96.04%  
97.13%  
1

Length: 14159  
Matches: 315  
Conservative: 2  
Mismatch: 5  
Indels: 6  
Gaps: 1

US-10-089-514-6 (1-322) x AF262220 (1-14159)

1 MetSerGIyPheProAGSerValValIGIyGIySerGIyAlaValIGIyMetPhe 20

|||||

10159 ATGAGCGGCTTCCCGGAGCGTCGTCGTCGGCGGAGCGGAGCGGCGGCGCATGTC 10218  
Qy 21 AIGIYleuLeuAargIuAagiYSerArgThrLeuValAlaAspLeuValProProPro 40  
Db 10219 GCGGCGGTGCTGGGAGGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 10278  
Qy 41 GYARProAspAlaCysLeuValIGIYAspValThraLProGIYProGIYleuAlaA 60  
Db 10279 GGAAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 10338  
Qy 61 AAlaLeuArgAspAlaAspLeuValLeuLeuAlaValAlaAspValAlaLeuYsAla 80  
Db 10339 GCCCTCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 10398  
Qy 81 ValAlaProValThraArgLeuMetArgProGIYAlaLeuLeuAlaAspThrLeuSerVal 100  
Db 10399 GTGCGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 10458  
Qy 101 ArgThrGIYMetValAlaGIYleuAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 120  
Db 10459 CGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 10518  
Qy 121 AspProMetPheAlaProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 140  
Db 10519 AACCGATGTTGCGCGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 10578  
Qy 141 ArgAspGIYProGIYValThraAlaLeuLeuArgLeuValIGIYGIYGIYArgPro 160  
Db 10579 AGGAGCGGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 10638  
Qy 161 ValArgLeuThraAlaGIYLeuValAspArgThrThraAlaAlaAlaAlaAlaAlaAlaAla 180  
Db 10639 GAGCGGCTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 10698  
Qy 181 AAlaValLeuLeuSerPheGIYleuAlaLeuAlaArgLeuGIYValAspValArgAlaLeu 200  
Db 10699 GCGGTATCT 10758  
Qy 201 AAlaAlaThraAlaProProProProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 220  
Db 10759 GCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 10818  
Qy 221 GIYSerProGIYValYArgIYAspIleGIYAspSerAspProArgAlaAlaAlaAlaAla 240  
Db 10819 GCGAGCGGCGGAGGTGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10878  
Qy 241 ArgAlaLeuAlaGIYAlaLeuArgSerPheAlaAlaLeuValGIYAsp----- 256  
Db 10879 CGGCGCTCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 10938  
Qy 257 -----AspProAspArgAlaAspAlaProGIYArgAlaAspAlaProGIYAspProGIY 274  
Db 10939 GCCGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 10998  
Qy 275 GIYCysAspGIYAlaAlaYAsnLeuAspGIYValPheGIYLeuLeuArgArgLeuMetGIY 294  
Db 10999 GGAATGCGAGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 11058  
Qy 295 ProGIYleuAlaAlaAlaGIYAspAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 314  
Db 11059 CGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 11118  
Qy 315 AspAspGIYGIYGIYAspArg 322  
Db 11119 GACGAGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 11178  
RESULT 5  
A48324 A48324 888 bp DNA linear PAT 07-MAR-1997  
LOCUS  
DEFINITION Sequence 2 from Patent WO9601901.  
ACCESSION A48324  
VERSION A48324.1 GI:2302117  
KEYWORDS



SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
ORIGIN

Streptomyces pristinaespiralis  
Streptomyces pristinaespiralis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 888)  
Blanc, V., Thibaut, D., Bamas-Jacques, N., Blanche, F., Crouzet, J.,  
Barriere, J.-C., Debussche, L., Famechon, A., Paris, J., and  
Dutric-Rosset, G.  
STREPTOGRAMINS AND METHOD FOR PREPARING SAME BY MUTASYNTHESIS  
Patent: WO 9601901-A 2 25-JAN-1996;  
RHONE-POULENC RORER SA (FR)  
Other publication AU 2891295 960209  
Other publication FR 2722210 960112.  
Location/Qualifiers  
1..888  
/organism="Streptomyces pristinaespiralis"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:38300"

Alignment Scores:  
Pred. No.: 1,27e-20 Length: 888  
Score: 661.50 Matches: 155  
Percent Similarity: 60.38% Conservative: 34  
Best Local Similarity: 49.52% Mismatches: 101  
Query Match: 40.46% Indels: 23  
DB: 6 Gaps: 4

US-10-089-514-6 (1-322) x A48324 (1-888)

QY 2 SerGlyPheProArgSerValValValGlyGlySerGlyAlaValGlyGlyMetPheAla 21  
DB 13 TCGGTGTTCCGGCGCTTGTGTGTGTGGCGGGCGCGTGGCGCCGATGTTACAG 72  
QY 22 GlyLeuLeuArgGluAlaGlySerArgThyLeuValAlaAspLeuValProProProGly 41  
DB 73 CACTGGCTGTGGCTTGGCGGGGTGGCGCGTACCTGGCTGAC--GTGGCGGGGGCGGT 129  
QY 42 ArgProAspAlaCysLeuVal-----GlyAspValThyAlaProGlyProGluLeuAla 59  
DB 130 GCGGCGGACGGGGTGGCGGGTGTGGCGCGGTGATGTGGCGGGCGGGCGGAGCGGCTC 189  
QY 60 AlaAlaLeuArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuLys 79  
DB 190 GCGGCGCTGGCGGGCGGCGGACGTGTGTGTGTGGCGGGTGGCGGACCGGTGGCGGAG 249  
QY 80 AlaValAlaProValThyArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSer 99  
DB 250 GCGGTGAGTGTGTGGCGGGGTGATGCGGGCGGTGTGTGTGTGTGTGTGTGTGTGTG 309  
QY 100 ValArgThyGlyMetAlaAlaGluLeuAlaAlaHisAlaProGlyValGlnHisValGly 119  
DB 310 GTCAAGAGCCGGATCCGCGGGCGGCTCCGTGAGCGGGCGGGCGGTGCAGGGCGGTGGG 369  
QY 120 LeuAsnProMetPheAlaProAlaAlaGlyMetThyGlyArgProValAlaAlaValAla 139  
DB 370 CTGAACCCGATGTTGTGGCGCCCTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 429  
QY 140 ThyArgAspGlyProGlyValThyAlaLeuLeuArgLeuValGluGlyGlyGlyArg 159  
DB 430 GTCAACGACGGGCGCGGTGTGGCGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 489  
QY 160 ProValArgLeuThyAlaGluGluHisAspArgThyThyAlaAlaThyGlnAlaLeuThr 179  
DB 490 GTGTGAGATGCGCGGGCGGCGGACGACGACGACGACGACGACGACGACGACGACGACG 549  
QY 180 HisAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValAlaGln 199  
DB 550 CATGCGCGGTG 609  
QY 200 LeuAlaAlaThyAlaProProProHisGlnValLeuLeuAlaLeuLeuAlaGlyValLeu 219  
DB 610 CTGGGAGACGTGCGCGGGCGGCGCATGTGGCATGTGTGTGTGTGTGTGTGTGTGTGTGTG 669

QY 220 GlyGlySerProGluValTyGlyAspIleGlnArgSerAspProArgHisAlaSerAla 239  
DB 670 GCGGCGACCCGAGGTGTATTTTCAGATCCAGGCGGCCAACCCGCGCGCGCGCGCGG 729  
QY 240 ArgAlaGluAlaLeuAlaGluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAsp 259  
DB 730 CCGACGGCCCTGCGGCGCGCGCGCTG-----GTGCGGCTGTGGCGGCGCGCGCTGAG 777  
QY 260 ArgAlaAspAlaProGlyValArgAlaAspAlaProGlyHisProGlyGlyCysAspArgAla 279  
DB 778 AGGCGCGACGAGAGACGTTCCG----- 801  
QY 280 GlyAsnLeuAspGlyValPheGlyGlyLeuAlaArgArgLeuMetGlyProGluLeuAlaAla 299  
DB 802 -----GCCCTGTCCCGAAGCTGCCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 849  
QY 300 GlyGlnAspHisCysGlnGluLeuPheArgThrLeuHis 312  
DB 850 CTGGAACGGCTGTGGCGCGCGATGTTCAACCCGCTGCAC 888

RESULT 6  
AR198354  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

AR198354  
Sequence 2 from patent US 6352839.  
AR198354  
AR198354.1 GI:20248203  
Unknown.  
Unclassified.  
1 (bases 1 to 888)  
Blanc, V., Thibaut, D., Bamas-Jacques, N., Blanche, F., Crouzet, J.,  
Barriere, J.-C., Debussche, L., Famechon, A., Paris, J.-M. and  
Dutric-Rosset, G.  
STREPTOGRAMINS FOR PREPARING SAME BY MUTASYNTHESIS  
Patent: US 6352839-A 2 05-MAR-2002;  
Location/Qualifiers  
1..888  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN

Alignment Scores:  
Pred. No.: 1,27e-20 Length: 888  
Score: 661.50 Matches: 155  
Percent Similarity: 60.38% Conservative: 34  
Best Local Similarity: 49.52% Mismatches: 101  
Query Match: 40.46% Indels: 23  
DB: 6 Gaps: 4

US-10-089-514-6 (1-322) x AR198354 (1-888)

QY 2 SerGlyPheProArgSerValValValGlyGlySerGlyAlaValGlyGlyMetPheAla 21  
DB 13 TCGGTGTTCCGGCGCTTGTGTGTGTGGCGGGCGGTCGCGTGTGTGTGTGTGTGTGTG 72  
QY 22 GlyLeuLeuArgGluAlaGlySerArgThyLeuValAlaAspLeuValProProProGly 41  
DB 73 CACTGGCTGTGGCTTGGCGGGGTGGCGGTGACCTGGCTGAC--GTGGCGGGGGCGGT 129  
QY 42 ArgProAspAlaCysLeuVal-----GlyAspValThyAlaProGlyProGluLeuAla 59  
DB 130 GCGGCGGACGGGGTGGCGGGTGTGGCGGTATGTGGCGGGCGGGCGGAGCGGCGGCTC 189  
QY 160 AlaAlaLeuArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuLys 79  
DB 190 GCGGCGCTGTGGCGGACGATG 249  
QY 80 AlaValAlaProValThyArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSer 99  
DB 250 GCGGTGAGTGTGTGGCGGGGTGATGCGGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTG 309



QY	100	ValaTgThgTgMeTAlaAlaGluLeuValaAlaAhiAaIaProGlyValaGlnHiAValGly	113
Db	310	GTCAAGACCGGATCGCGGGCGGCTGCGTGAAGCGCGCGGCGTCAAGCGGATGGG	369
QY	120	LeuAaNPromEPhaAlaProAlaAlaGlyMeTThrGlyArgProValAlaAlaValVal	139
Db	370	CTGAACCCGATTTGGCCCCCTTCGCTGGGTCTTCAAGGGAGGCGCGGTGGGTTGG	429
QY	140	ThraGaaPglyProGlyValThraAlaLeuLeuArgLeuValaGluGlyValGlyValArg	159
Db	430	GTCAACGACGGGCGCGGTGTGGGGCCCTGGTGAAGCTGTGTGGCGGGTGGGGCGCG	489
QY	160	ProValaTgLeuThraAlaGluGlnHiAaPaTgThrThraAlaThGlnAlaLeuThr	179
Db	490	GTGTGTGAAGATCCCGCGCGGCGGACAGACAGCTGACCGCGCGCACAGCGCCGACG	549
QY	180	HisAlaValLeuLeuSerPheGlyLeuAlaLeuAlaAlaGluGlyValaAspValArgAla	199
Db	550	CATGCGCGGATCTGACCTTCGGGCTGGGCTGGGTGAAGCTGTGATACGTGGGGCG	609
QY	200	LeuAlaAlaThraAlaProProPheGlnValaLeuLeuAlaLeuAlaAlaArgValLeu	219
Db	610	CTGCGGGACAGTGGCCCGCGCGGATCTGGGAGTCTGGCGCTGTGGCCGATCGCC	669
QY	220	GlyGlySerProGlnValaTgTgValaPheTgLeuArgSerAspProArgAlaAlaSerAla	239
Db	670	GGCGGGACCGCGGAGTGTATTTCAGATCCAGCGCGCCAAACCCGCGCGCGGCGCGG	729
QY	240	ArgArgAlaLeuAlaGluAlaLeuArgSerPheAlaAlaLeuValaGlyAspAspProAsp	259
Db	730	CGGCGGGCTGTGGCGCGCGCTG-----GTGGCGTGTGGGCGAGGCGCTGCAG	777
QY	260	ArgAlaAspAlaProGlyArgAlaAspAlaProGlyHiAProGlyGlyAspAlaAla	279
Db	778	AGGGCGGACGAGGAACCTGCC-----801	
QY	280	GlyAsnLeuAspGlyValaPheGlyGlyLeuArgArgLeuMetGlyProGlnLeuAlaAla	299
Db	802	-----GCCCTGTTCGCCGAATCGCGGTGTGGGCGAGCACCGCGCGGAG	849
QY	300	GlyGlnAspHisCysGlnGlyLeuPheArgThrLeuHiA	312
Db	850	CTGGAACGGCTGTGCGCGCGGATGTTCACCGCCCTGCAC	888
RESULT 7			
LOCUS	A48323	2888 bp	DNA linear PAT 07-MAR-1997
DEFINITION	Sequence 1 from Patent WO9601901.		
ACCESSION	A48323		
VERSION	A48323.1 GI:2302116		
KEYWORDS	.		
SOURCE	Streptomyces pristinaeepiralis		
ORGANISM	Streptomyces pristinaeepiralis		
REFERENCE	Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;		
AUTHORS	Streptomyceinae; Streptomycetaceae; Streptomyces.		
	1 (bases 1 to 2888)		
	Blanc, V., Thibaut, D., Bamae-Jacques, N., Blanche, F., Crouzet, J.,		
	Barrere, J., Debusche, L., Fanechon, A., Patris, J. and		
	Dutric-Rossec, G.		
TITLE	STREPTOGRAMINES AND METHOD FOR PREPARING SAME BY MUTASYNTHESIS		
JOURNAL	Patent: WO 9601901-A 1 25-JAN-1996;		
	RHONE-POULENC RORER SA (FR)		
COMMENT	Other publication AU 2691295 960209		
	Other publication FR 2722210 960112.		
FEATURES	Location/Qualifiers		
Source	1..2888		
	/organism="Streptomyces pristinaeepiralis"		
	/mol_type="unassigned DNA"		
ORIGIN	/db_xref="taxon:38300"		
Alignment Scores: 3.69e-20 Length: 2888			

[illegible]



[illegible][illegible]











WTSAGVAGIDIALALAEEDFGKALIDVARILWELKPGGOEFSAPLALOSTSK  
FADHAWMTNNIGDGLSVATLATRAMSRSPFYWRHSHAGITTPAKAYERRRLBAABH  
LVYSTLSLKICIASRCGFCGEDTLRNSFFRASSTPQDIRKHKFTTP"

10003..10784  
/locus\_tag="plu3545"  
10003..10008  
/locus\_tag="plu3545"  
10017..10784  
/locus\_tag="plu3545"  
/function="Unknown"  
/note="unnamed protein product; Some similarities with  
unknown protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAE15918"  
/db\_xref="GI:36786853"  
/translation="MKRTSGEHPALSSLINSSGKPLPAFVIRPKGTIDMIGPAT  
VLVGWGLNLEIYKWRDLTPRTLAIISPSRTEQLNKVDILCVPTGNVALLAE  
LDMLDMLHFKGSGKSWTSSVCTSIILGAGLLKGYOATHMSEIDDKPDAIPVAK  
RVVDIRRVRVGGGTGADIDGELLRLLEIGEEAQAQVQLAMQSYSEPPFNSGMPEDP  
KEITHATLOAIFSNVEKQGTDEYLEILIKSAKRISDIY"  
complement(11533..12540)  
/gene="ISPlu15A"  
/locus\_tag="plu3546"  
complement(11533..12540)  
/gene="ISPlu15A"  
/locus\_tag="plu3546"  
/function="Transposase"  
codon\_start=1  
/transl\_table=1  
/transl\_table=1  
/protein\_id="CAE15919"  
/db\_xref="GI:36786854"  
/translation="MKRSHKTPDPAIFPKPLSHIDTARDPLFELPATTLVACDLDLT  
LRSSGFIIDNLIRKSHDLYSLKTGSESVYCVIHOSPDKMAFLRMYISAR  
MORLBOGHKKLPLVPLFVFNKRIPIPYMSTNMPDCCDASLAEIYSKFLPVDTVT  
GIDPDELHTRALAEIVKHLIRQDMALQOELTMLFADYYVEELKSMNLVITLL  
IGVDELADGELTRQALAEPPYEEVLTIAIQKLIKNGHOGKLEGLQCODAREEGLOE

[illegible]

US-10-089-514-6 (1-322) X BX571871 (1-349944)

Oy	8	ValValValGlyGlySerGlyAlaValGlyGlyMetGluAlaGlyLeuLeuArgGluAla	27
Db	30323	GTCAATCCCTGGATGGGCAAGGCGGCTAATGGTCTTATTAAGCGCGCTTTTACTCAATAT	30266
Oy	28	GlySerArgThrLeuValAlaAspLeuValProProGluArg-----	42
Db	30263	GGTATATATGTAATACAGTGTGGACAAACGACGAGGAGTGGCTTCTACTATCAT	30206
Oy	43	ProAspAlaCysLeuValGlyAspValThrLeuProGlyProGluLeuAlaAlaLeu	62
Db	30203	TTCATTAATCTGACGATT---GATATCTTAAACCGGCTGCTATGTGGAGCGGTTTGG	30144
Oy	63	ArgAspAlaAspLeuValLeuLeuAlaValAlaHisGlyAspValAlaLeuValAlaAla	82
Db	30146	AGTAGGGCAAGGCGGCTGCTTGGCTTACCGGAAACCGTCGGACAAAGCGTTGGCT	30088
Oy	83	ProValThrValGlyMetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThr	102
Db	30086	TGGGTGTATCTGCTATGATGGATGATGATACAGATCCGTGTCAACCTGGTACGTTCAAGGCC	30027
Oy	103	GlyMetAlaAlaGlyLeuAlaAlaHisAlaAlaProGlyValGlnHisValGlyLeuAspPro	122
Db	30026	CCCTTTTATTTGCAATTAAGAGTGGCGGCGCGCGGAGCGCTTTCATGGATGGTTAACCGG	29966

[illegible]

```

RESULT 11
LOCUS      AX770909                      349980 bp    DNA       linear    PAT 02-JUL-2003
DEFINITION Sequence 40 from Patent WO02094867.
ACCESSION  AX770909
VERSION     AX770909.1  GI:32438073
KEYWORDS

SOURCE
ORGANISM   Photorhabdus luminescens
            Photorhabdus luminescens
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Photorhabdus.
REFERENCE  1
AUTHORS    Duchaud,E., Taourit,S., Glaeser,P., Frangeul,L., Kunst,F.,
            Danchin,A. and Buchrieser,C.
TITLE       Sequence of the Photorhabdus luminescens strain TT01 genome and
            uses
JOURNAL     Patent : WO 02094867-A 40 28-NOV-2002;
            INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
            SCIENTIFIQUE (CNRS) (FR)
FEATURES
    source              Location/Qualifiers
                        1..349980
                        /organism="Photorhabdus luminescens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:29488"
                        /note="Sequence length too big . Splitted into 2
                        sequences. ~seq 00040: 1 to 34980~seq 10790: 300001 to
                        649980~seq 10791: 600001 to 656798"

```

Alignment Scores:	
Pred. No.:	2,95e-06
Score:	382.00
Percent Similarity:	47.94%
	Length: 349980
	Matches: 101
	Conservative: 50

```
sequences.-seq 00040: 1 to 349980-seq 10790: 300001 to
649980~seq 10791: 600001 to 696798"
```

Alignment Scores:	
Pred. No.:	2,95e-06
Score:	382.00
Percent Similarity:	47.94%
	Length: 349980
	Matches: 101
	Conservative: 50



Best Local Similarity: 32.06% Mismatches: 126  
Query Match: 23.36% Indels: 38  
DB: 6 Gaps: 5

US-10-089-514-6 (1-322) x AK770909 (1-349980)

```
QY      8 ValValaValGlyGlySerGlyValaValaGlyGlyMetPheAlaGlyLeuLeuArgGluAla 27
      115977 GTCATCTTGGTGGGAGGCGCATATGCTTTCTTTATTAAGCGCCCTTTTACTCAATAT 116036

QY      28 GlySerArgThrLeuValValaAspLeuValProProProGlyArg-----42
      116037 GGCATATATTGTATACAGTGTGTCCAAACGACCGAGGAGTGGTCTCATTAATCATAT 116096

QY      43 ProAspAlaCysLeuValGlyAspValThraProGlyProGlyLeuAlaAlaLeu 62
      116097 TCTCATTAATCTGTAGAT---GATATCTTAAACCCGGCTGCTATGTGGAGCGGTTTG 116153

QY      63 ArgAspAlaAspLeuValLeuLeuAlaValaHisGluAspValaLeuLysAlaValaAla 82
      116154 AGTGGGCGCAAGGCGGTCCTTTGGCTTACCGGAGCGCTCCCATACAAAGCGTTGCC 116213

QY      83 ProValThraGlyLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThr 102
      116214 TGGGTGTATCTCTGCTATGGTGTATGATGATACAGATGCTGTCAACCTGTTCAAGTTCA 116273

QY      103 GlyMetAlaAlaGlyLeuValaAlaHisAlaProGlyValaGlnHisValGlyLeuAsnPro 122
      116274 CCTTTTATTCTGTCATTAAGAGCTGCCGCCCGGCGAGCGCTTTCATGTGTTAAACCG 116333

QY      123 MetPheAlaProAlaAlaGlyMetThrGlyArgProValaAlaValaValaThraArgAsp 142
      116334 ATGTTCTCGCATCTTTGGCCGAGCAAAAAGGCGCATGTCAGTCAATCTTCGATTAACGT 116393

QY      143 GlyProGlyValaThraLeuLeuArgLeuValGluGlyGlyGlyArgProValArg 162
      116394 CATGTTAGTGAACATTGATGTGAGCGGGTATATGATGAGGCTGATATGCTATCAGTAGG 116453

QY      163 LeuThraAlaGluGlnHisAspArgThrThraAlaAlaThrGlnAlaLeuThrHisAlaVal 182
      116454 ATGACCCCGAGAGAACACGATCAGTTATAGCCCTTTGTACAGCTTTGCCCATGCGCG 116513

QY      183 LeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValaAspValaArgAlaLeuAlaAla 202
      116514 ATACTATTTTGGCTTGGCTTGGCGCAAAAGTCCGTAGATATTCGTAACCTTGGCAGCT 116573

QY      203 ThraAlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValaLeuGlyGlySer 222
      116574 ATTGCCGACCGCCCATGCGTACCATGCGTGGCTGTATATCAAGATCTTAACGCAATCCA 116633

QY      223 ProGlyValaGlyGlyAspIleGlnArgSerAsnProArgAlaAlaSerAlaArgAla 242
      116634 ATGAAAGTGTACTGGGACGCTCAATATGAAATCCGATGCTGCCAAGCGCTCAGAG 116693

QY      243 LeuAlaGluAlaLeuArgSerPheAlaAlaLeuVal-----GlyAspAspProAspArg 260
      116694 TTCTCGATGTCATACACAGTGTAGATTGATGATGATGATGATGATGATGATGATGATGAT 116750

QY      261 AlaAspAlaProGlyArgAlaAspAlaProGlyHisProGlyGlyCysAspGlyAlaGly 280
      116750 -----TTTGGCATTAACCTCAGAGATATTCGCAACATCTGGGAGAACAG 116795

QY      281 AsnLeuAspGlyValaPheGly-----GluLeuArgArgLeuMetGlyProGly 296
      116751 -----TTTGGCATTAACCTCAGAGATATTCGCAACATCTGGGAGAACAG 116795

QY      297 LeuAlaAlaGlyGlnAspHisCysGlnGluLeuPheArgThrLeu 311
      116796 CTGTGATATCAGTCTGCTGCTGATTCGCAAGATATTTTTCACAGCTT 116840

DB      116796 CTGTGATATCAGTCTGCTGCTGATTCGCAAGATATTTTTCACAGCTT 116840
```

RESULT 12  
AE010371  
LOCUS AE010371 14192 bp DNA linear BCT 01-AUG-2002

```
DEFINITION Methanopyrus kandleri AV19 section 70 of 157 of the complete genome.
ACCESSION AE010371 AE009439
VERSION AE010371.1 GI:19887260
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Slesarev,A.I., Mezhevaya,K.V., Makarova,K.S., Polushin,N.N., Shcherbina,O.V., Shakhova,V.V., Belova,G.I., Aravind,L., Natale,D.A., Rogozin,I.B., Tatusov,R.L., Wolf,Y.I., Stetter,K.O., Malykh,A.G., Koonin,E.V. and Kozlovskiy,S.A.
TITLE The Complete Genome of the Hyperthermophile Methanopyrus kandleri AV19 and Monophyly of Archaeal Methanogens
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 14192)
AUTHORS Slesarev,A.I., Mezhevaya,K.V., Makarova,K.S., Polushin,N.N., Shcherbina,O.V., Shakhova,V.V., Belova,G.I., Aravind,L., Natale,D.A., Rogozin,I.B., Tatusov,R.L., Wolf,Y.I., Stetter,K.O., Malykh,A.G., Koonin,E.V. and Kozlovskiy,S.A.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-2002) Fidelity Systems, Inc., Gaithersburg, MD 20879

FEATURES
source
Location/Qualifiers
1..14192
/organism="Methanopyrus kandleri AV19"
/mol_type="genomic DNA"
/strain="AV19"
/db_xref="taxon:190192"
71..1321
/gene="MK0793"
71..1321
/codon_start=1
/gene="MK0793"
/transl_table=11
/product="Unclassified conserved protein"
/protein_id="AA02007.1"
/db_xref="GI:19887261"
/transl_table="MITSUUVSVSREITKGVSEIRVGIVIHGPTVIDGMARIRSLKILGVRKAGCTGTGVALIDGLDSVYFPRDKLSECLSEINEMADIVLITNPKSKRBSGLPAEOLVSRAGVDSLVOAERPGEDGIVLMPEDARTVAEHLREGLKIT EAVRTVTEKGGIGSRVACVEPDRILVNGITVGASRNVLEVPDSDGLVIRGGRIKPEGVRLGVDPERVVVKTDRRLRTERBERKRVSGPERIHRVILLVDARKV EDRMSDAVSVGDDTTCVCAELGRLGVWVGLVLDPDGDVNDOTRESLASEELA ALLVCERDDDAKILVRGPFRRERKRVLDPTVGEIVVEVCEKVKCYHHAKE TSA"
2581..3564
/gene="MK0795"
2581..3564
/gene="MK0795"
/codon_start=1
/transl_table=11
/product="Selenophosphate synthetase"
/protein_id="AA02009.1"
```



```

/db_xref="GI:19887263"
/translation="MKRLRLARLRSFEGVTRKGPVKTVEGLEPLDVTYFEGEVIAD
VGDAAAVKVDIILLIADGIMGKLIIDRPMPAGCAVIVNNVDIAMLGMPGVGVN
VISTDVNVCREILBEGMRGAKMKESTPVLGHTHPPTPTAYDAALVGTDEHLVLS
STBEGDLIVFDLNGRPYPERPLAMDTTGMKQKAVGAVASGLVAGKDV
SNGGLVGLAMLEASGCLGAEVWLDSIPKPEVDPMYTLKMPGKGFVYAVDSERDV
RAIRRLVSDARLDEVSIVGEVTEDGVKVKSEGSARVFPFERDRIILGVSPR"
3561. .5096
/gene="MK0796"
3561. .5096
/gene="MK0796"
/codon_start=1
/transl_table=11
/product="Predicted cyclophilin type peptidyl-prolyl
isomerase"
/protein_id="AA02010.1"
/db_xref="GI:19887264"
/translation="MKRVYVDGEVPDVEEATVADALEAGVSPEDVTIAVFGKQKV
ERETDRLIMLETGEDELSTVAVEDERMSVCEELPGASVSWTRDEVLGFGVDSD
LEFHTRGVVEPPYTAIILPTNDSEAYELIKRRMAVEYICTDIHGRTAGRELVD
ALRGSRVHVHEPVRATERVVSRYTLDDGLEAGRIITRVEIELEKAPVSAEHL
NTLEMEGRRIKFRDTFTSIEPRPDIPEENVDMREKVTVRRKGVDSVATVY
RDRTPVESHNVAVVRKMELLDVAAGDRLVETDPPRVNFGVLTVDGARLAEF
DVELEVNGDGVVDQEPRETLNVLKERKVVVEVDEVIETELYEDDAPRSVEYFR
RTYKMLDRPVGRLEKHAAYADLGMIVFEGNEKLGKLLPENNPXDRVEAGVIGVTVQA
KPHAGLIGVRLDESEYGPTEFEGTNVIGRVVEGLGRLEMDQSDMGRVTVVREVR
GR"
5093. .6400
/gene="MK0797"
5093. .6400
/gene="MK0797"
/codon_start=1
/transl_table=11
/product="Predicted Fe-S oxidoreductase"
/protein_id="AA02011.1"
/db_xref="GI:19887265"
/translation="MKVLVSPETTYTGAIVAVGCEBHSHTILRRSPDEALKRAD
VILSHHTLHLDEDLLEIARQAHFGKPVIVGSGVQPELVLERFPATYAKQEA
ENGLPRLQTLBEGDFEDVEGIALRDGEIVDTGMPRPADIDGSPRLKYPRLGQD
VRGANYIETHRCGPACTFCQVPEFFGRVRWKPEAVALEVRELTVRGARFRFALSG
GVTYTYGDEBEDVELLKRLLADLIGRENSAPVRADLNLNRLLEAIRYTTGMITLG
IBSGDRILARAKGKITVDVCEAVELARVAVAGSFIVGPGETEDLEATEELL
TELNLDDVFINLAEPITGTELGLTELPEEIPVLRPGEQLTEAGDALQLQTLAF
TTLRSVPLTDDVFHEALNIRDERKVLRTFRLGAGCCGT"
6397. .7659
/gene="tyra_2"
/note="synonym: MK0798"
6397. .7659
/gene="tyra_2"
/codon_start=1
/transl_table=11
/product="Prephenate dehydrogenase"
/protein_id="AA02012.1"
/db_xref="GI:19887266"
/translation="MRILILGSGTGMGRLLIARLRDGHVVITGSNPHTRARVAREL
DVEAATPNDAKDAVVVVSPISTEDVIRVAAHVVEGSLITVTYSKARPVAM
LEHAPDVVLGTHPLFGPTVPSLRQTVILITERSGPTVRVRVLERKGRVET
TPEHRTMAVVOCLTHAVLLAGAAIGRFLPELIDIEVNASPVYRLMDVGRVLAG
ODPRLVAEIOAFNPYGDDEAREBELRLARFHEHAHNMALTEXIASRRELGELELDE
ACORRTDKLSYLADELILQSGIEAILDVIDVNAALGEGVTGDRDTRIRDGSELPRD
RYKILPRCPKELAHFDGEVTVLELTGADGDVAAVVDALGRRLRLARBERGERVAV
EVEYTKESIEGLRKLRLALGLEVRAR"
6397. .7659
/gene="MK0799"
6397. .7659
/gene="MK0799"
/complement(7707. .9713)
/translation="MK0799"
6397. .7659
/gene="MK0799"
/codon_start=1
/transl_table=11
/product="Coenzyme F420-reducing hydrogenase, beta subunit
fused to oxidoreductase related to nitrite reductase"
/db_xref="GI:19887267"
/translation="MRSIVTSEHGHLVTFEPEMKLTBEVVEPGACALGCTVACIRG

```

```

GIIETDEGRLTEECARKGTGNTVCPRVDTAAVYHLGRLIAGRDYERPLTGYRRVAV
GVAADSDRLREIGOGCAVATALALAEGLADAVGVAGSAMPKCTVVEDEPEVK
DIAGSKYTRVGVLEALAEADAGIRVIALIGLPCVONGAKIOHPEITVAKGABALNI
DSSPAKRLPEVATATIGLPTKTRPEVEGLYKLIAREKGVDIEDERPITSGKLARVLSG
GATKEVDLAEFEEDALPEGRCINDPTARLADSVSGVTPREGVTTLLISEGEEVLE
GAVEAGYLVLRVANGSDVRRLAKLCKDAMKEAKNLEKRRKPPFWGVDGCVTLR
AGTEFVRLKVPFGVDDPDI FRALISGLEGYRAKFPDROOLEIHGIIPAEIPEVV
ERLEGTNLVNGSGSLVRTIMACPGKDNSSGILNTEELARLKEELAEPEPYKFI
ALSGTCNSGVRPOHHDIFRAGAVRPGVDEPKTCGQCVADCKVADRILITVGGAAV
ADTYKRCVYCGKCTINVCPEKARYAKESGIYVIGKGRREPEBGRALVDFADPSIP
TIAHTVLETRYLAEBPQERLADTIRKGLRQFTEALIRVLISQ"
9775. .10158
/gene="MK0800"
9775. .10158
/gene="MK0800"
/codon_start=1
/transl_table=11
/product="Predicted transcriptional regulator containing a
with DNA-binding domain"
/protein_id="AA02014.1"
/db_xref="GI:19887268"
/translation="MLADVPFVTRLVNTVDVLRGVGLRDEVDYTFALLERSEA
TVYDLAEELDRDRTVQKLKSLVYAGLVTRRKETPRGGFVYVYKAVPBEARKIVL
RALDEVEAVAKDRLERADVPRSGPE"
10155. .10766
/gene="MK0801"
10155. .10766
/gene="MK0801"
/codon_start=1
/transl_table=11
/product="Oxidoreductase related to nitrite reductase"
/protein_id="AA02015.1"
/db_xref="GI:19887269"
/translation="MIOSPDTYVVRLLAPGPCVAKELIARTVAELRYASGEVHITV
RGLLEIPEVPSKLDLRLKRLGLEPDSGTBPRVQVCCGTRCANALVDPVLA
RKLHEFELVWYPAVKILVSGCPGCTGSPENDLGLVAVGDEMLVGGYVVALP
BODVDVAVELTLEWYASBAPPGTRLHFRFTDVSELDVLERL"
10720. .11985
/complement(10720. .11985)
/gene="MK0802"
10720. .11985
/gene="MK0802"
/codon_start=1
/transl_table=11
/product="Unclassified protein"
/protein_id="AA02016.1"
/db_xref="GI:19887270"
/translation="MLPMLFTLLLAGPSQAELGPAVDPGEGTASNPALAQSDHA
LIAMKSPGELGIAVAKADGRVAVKGIPTPVASYPAAAPMRGERDLATFVLTVPLMT
BGRPIDVDIVESDVNCAVILMTGLSVRGDPAVAVVDSTFALVYRCEGLRVRGN
VNGVWRMTDEBQGVPELKLSPGPERERSVSVAGRPPTYLLVAYETPAGREVAVN
LTPDPAHPTIDGAPIPVPGTRPVLLDNVLAIVGADGVHAALLVKGTRAKIASDVKADA
TTPVRAIVSGVDRYLIYVGSSELRLAAEVRIGTRGVVRBETLIKEPYSPTGIGV
RTTRESLVEAGVILAFYADGEPFVHVEYSBGLPLTQHRRRRPVIVPLLAASIGA

```

## Alignment Scores:

Pred. No.:	5,52e-05	Length:	14192
Score:	323.00	Matches:	93
Percent Similarity:	52.55%	Conservative:	41
Best Local Similarity:	36.47%	Mismatches:	104
Query Match:	19.76%	Indels:	18
DB:	1	Gaps:	8

US-10-089-514-6 (1-322) x AB010371 (1-14192)

Qy	6	ArgSerValValValGlyGlySerGlyAlaValAGlyGlyMetPheAlaGlyLeuLeuArg	25
Db	6400	AGGATCGCATACTCGGCGGTACCGGCGCCGCTTATGCGCCGCGAGCGTCCGT	6459
Qy	26	GlualAGly-SerArg-----ThrLeuValValAspLeuValProProGlyArgPr	43
Db	6460	GACGACGGTCATGACGTACTTATCACTGCTTACCCCTCACACCGCCGACGCGTCCG	6519
Qy	43	oksAspIaCySLeuValAGlyAspValThrAlaProGlyProGluLeuAlaAlaLeuLys	63



```
Db      6520 CGGAGAGCTC-----GACGTCGAGGCT---GCTCCGACCAACGTCGATCCGCCCA 6566
Qy      63  GAAPAAAPLeuValLeuLeuAlaValHisGluAspValAlaLeuLysAlaValAlaPr 83
      6567 AGATCCGACGACTGTGTGTGTTCCGTCCTCCCACTCCGTAAGAGAGCTCATCCGTGA 6626
Qy      83  oValThrArgLeuMetArgProGluAlaLeuLeuAlaAspThrLeuSerValArgThrG 103
      6627 GGTTCGACACACAGCTTCCTGAGGGTTCCCTCTCACGACGTAAGCTCGGTCAAGGTTG 6686
Qy      103 yMetAlaAlaGluLeuAlaAlaHisAlaPro---GlyValGlnHisValGlyLeuSerPr 122
      6687 TCCCGTTCCGCTAGTATGTCGAGACGCTCCGAGATGTTGTAAGTCTCTCGAAGTCAACC 6746
Qy      122 oMetPheAlaProAlaAla---GlyMetThrGlyArgProValAlaAlaValAlaThr-- 140
      6747 CTTCTTCGGGCCCAACGCTCCCAAGTCTCCGCGGTCAAGCTCATCTCAAGCCCAACGGA 6806
Qy      141 -ArgAspGlyProGlyValThrAlaLeuLeuAlaGluGlyGlyValArgPr 160
      6807 GCGTAGAGGTCCCTGGAGCTCGCCGCGTTGCTGTTATCTGAGCGGAAAGCGCTCGAGT 6866
Qy      160 oValArgLeuThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHi 180
      6867 CGTTGAGACCAACCCCGAGGAGAACACGATCCACTATGCGCGTTCAGTGCCTCAACCCA 6926
Qy      180 sAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArg-----LeuGlyValAla 196
      6927 TCGTGTCTCTCTGCTGCTGACAGCCCGCCATAGGCGGTTCTCTCCAAAGTTCAGAGCTCGA 6986
Qy      196 pValArgAlaLeuAlaAlaThrAlaProProProHisGlnValLeuLeuAlaLeuAla 216
      6987 CATCGAAGAGCTGCGCCGAGC-----CCCGTATGCGCTACTATGAGCGTCTGAGG 7037
Qy      216 aaArgValLeuGlyGlySerProGluValTyrGlyAspIleGlnArgSerAsnProArgAl 236
      7038 TCTATATCGCGGCGAGGATCAGCTTTTACGCGGAGATCAAGCCTTCAACCCCTTACG 7097
Qy      236 aaIAsSerAlaArgArgAlaLeuAlaGluAlaLeuArgSerPhe 250
      7098 TGACGAAGCCCGTGAAGAGCTCTCTCGCGCCCTCAAGCCTTC 7140
Db

RESULT 13
LOCUS   AE011181                      12180 bp    DNA       linear    BCT 03-APR-2002
DEFINITION Methanosarcina acetivorans str. C2A, section 526 of 534 of the
complete genome.
ACCESSION AE011181 AE010299
VERSION   AE011181.1 GI:19918739
KEYWORDS
SOURCE    Methanosarcina acetivorans C2A
           Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
           Methanosarcinaceae; Methanosarcina.
REFERENCE
AUTHORS  Galagan,J.E., Nussbaum,C., Roy,A., Smirnov,S., Arnoor,D., Brown,A.,
          Fitzhugh,W., Calvo,S., Engels,R., Edrington,N., Dearliano,K., Johnson,R.,
          Lincon,L., McEwan,P., McKernan,K., Talamas,D.E., Grahame,D.A.,
          Zimmer,A., Barber,R.D., Cann,I., Graham,J., Tittell,A., Ye,W.,
          Guse,A., Hedderich,R., Ingram-Smith,C., Kuetner,C.H.,
          Krzycki,J.A., Leigh,J.A., Li,W., Liu,J., Mukhopadhyay,B.,
          Reeve,J.N., Smith,K., Springer,T.A., Umayam,L.A., White,O.,
          White,R.H., de Macario,E.C., Ferry,J.G., Jarell,K.F., Jing,H.,
          Macario,A.J.U.L., Paulsen,I., Pritchett,M., Sowers,K.R.,
          Swanson,R.V., Zinder,S.H., Lander,E., Metcalf,W.W. and Birren,B.
          The Genome of M. acetivorans Reveals Extensive Metabolic and
          Physiological Diversity
          Genome Res. 12 (4), 532-542 (2002)
JOURNAL  Genome Res. 12 (4), 532-542 (2002)
MEDLINE  21929760
PUBMED   11932238
REFERENCE 2 (bases 1 to 12180)
```

```
AUTHORS  Birren,B.
TITLE    Direct Submmission
JOURNAL  Submitted (20-MAR-2002) Center for Genome Research, Whitehead
          Institute, Nine Cambridge Center, Cambridge, MA 02141, USA
FEATURES
source   Location/Qualifiers
          1..12180
           /organism="Methanosarcina acetivorans C2A"
           /mol_type="genomic DNA"
           /strain="C2A"
           /db_xref="taxon:188937"
           complement(541..4614)
           /gene="MA4589"
           complement(541..4614)
           /gene="MA4589"
           /codon_start=1
           /transl_table=11
           /product="cell surface protein"
           /protein_id="AA07928.1"
           /db_xref="GI:19918740"
           /translation="MMNIRSKYAVYIFPLLILPFVGTASALPTVAHDRVQGEVYAS
           TAMDSHSTNSFDVNGSVFARYVGVVAGSTPSISATFNGHAFETNPBYISGM
           GVTWLPENVTDVQDEMTNATIDSHNGDGRQVGSITLVYLENKEPVEVIAEGL
           DMVITGSGSTTYFNDNTIDLDVDASLXSTILITGYSSESLNGESLSCAVDSGATP
           NTRKDNVSDSLTAENQITIKVGDGYTSVAVHGLSITTKLPDLVPTSLTBSYVTPGT
           VMTWATIENTDKDPSFNVS.LIADGTAVDDVGVGIAAGSTVDLQMTPEGADSY
           SLTVVDDPNSVVESEETNELTVLVGSAVSVPAANFTAVPTSGSILPLVFPDOST
           NSPTSMFMDFGDENSESTSONPSHTVTSVGNHTSLVTNNAEGDSEVKDYITVEVIT
           PVAARTANDTSGAAPFTVQFPDQSANLPAEMFMDGDPDGNSTREONPAHTVITAGKYV
           KITTNANAGNDESEKTDITATIPVPAATPATPTGGAALTYTPTDTSNNTEBMLMD
           FDDGNSTQDNPIHTYTSAGNTVAFATNSGSGSELKATVITVTAEGKGLADTA
           WPKFHYDLGNTQSPYTGTPQANVVMVYTTTGEMLGOGAVIGSDGTWYIIGRSGTLY
           AINPGLTKGTSIGTSKSFPTPAISDGTIYGANDKIYVAINPDGSIKMSYPVGTLS
           VISSPAISADGTIYATSTPALPKPDGTLKMSAVSGVYVFSFPAIGSDGTIYIGT
           YGSLAINADGTLTAMTYSIPAPNNYITDTPALASDGTITGECNNANIALNPDLTK
           WTVSGTSSQAKSKSPAGSDGTITRSGNNOYFALIPDGLTKSSRAGSGCQTPT
           TGSDDGTINDGSKTLYAFNPBDGVTKMSYADGNVYASPAISDGLTFKGGYTLVA
           FPDVAPVSNFTADPTSGTAPALVNTFTDSTGSPSWFMDFGANSTEQNPHTVSA
           GNYTTNLIVTNSAGDPEVKTDYITVLPSSTPEPVAFTADVSGSAPLVNFTSD
           STGSPSLMLMDPGDGNSTREONPHTVYASGVYTNLVENDAGSDPELKTDTYIYVE
           ISGIVVTLYFPDESSVAENESTELSIIVASNPAGLSGNTLVAILDDPAVAIEVIEY
           PFWALITERSTIPGTSIYKTYDLEDSIKEEGGEVVLATUTYSGEKSSANISITSEY
           LEDDSGDSLBPALTLGITLBTLYLSLPPDETA PKLDGDLGYEDLTGNGEFSVDIVA
           YPHNMDWIENNPPVEYFDFNGNGRIDFDVDMFAMI"
           complement(5828..6397)
           /gene="guaa"
           /note="synonym: MA4590"
           complement(5828..6397)
           /gene="guaa"
           /codon_start=1
           /transl_table=11
           /product="GMP synthase (glutamine-hydrolyzing)"
           /protein_id="AA07929.1"
           /db_xref="GI:19918741"
           /translation="WRELKILVNNVGGFCHLIRAVBDLMDTKIIPNVETIEDILA
           BEPDLILSGGEMERAGICPDVREIDIPILGILGQALALAGVHSHSKKGQYA
           EIEIIVIEDDLIRGLGPKITVWASHADEVALLPGLGILHARSDCIEAHMPPKPI
           YGVQWHPVSHKXGDELLTNFEFCDR"
           6911..7708
           /gene="deoc"
           /note="synonym: MA4591"
           6911..7708
           /gene="deoc"
           /codon_start=1
           /transl_table=11
           /product="deoxyribose-phosphate aldolase"
           /protein_id="AA07930.1"
           /db_xref="GI:19918742"
           /translation="MSEIGKIRIRLRANRESNNVYIIPMDGISDGPLEGILNITDT
           VNRVVEGGANAVIMQKGVKYGREGYDIGIIVHISSVYSPBCKKIKNVCPDEEV
           IYKMGADVSMHINSGSNTEADOLEMGLISRCTWKMPLAMVYPRCKKITPNPDDV
           NVAHARIGAEIGADVKTIVYTGDDPSFRDVRGCPVPVVIAGFKSTIDMELMIDV
           GAIEAGARGAALGRNVFOHRDPRVLRITACETIVHRRPVEALEQLK"
           7816..8958
           gene
```



gene "MA4592"  
7816..8958  
/gene="MA4592"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="AA07931.1"  
/db\_xref="GI:19918743"  
/translation="MKKSWIKADDEGMEQKDRITTGESGADCVLPNGVEKRR  
ELGNTVAARNDKSRADIVVKGREGGQTPLEIGSFEDINAATLMDKGVV  
GGYVVIKDKHYEFAEMGKICDYLVTGDMKVIPLNLIADLQHKVKIIFGVSA  
BEARLAFOQLTAGDGLDLSGNPOEIKDTIKARLESSEAEALAVTVREPLG  
DRVCVDCNLMGEGEMLIGSAGMFLVNSDDSDSYVAAPRPNVAGVHSYIKG  
ETRYLSELADGPRTVDSKGRREGIVGRVKIESRPLMLEKAPDRUTLALQNA  
ETIKLVGKDGTPTSVALEKDEVLVLEBGRHFGKITEITIEK"  
9118..9846  
/gene="arod"  
/note="synonym: MA4593"  
9118..9846  
/gene="arod"  
/codon\_start=1  
/transl\_table=1  
/product="3-dehydroquininate dehydratase"  
/protein\_id="AA07932.1"  
/db\_xref="GI:19918744"  
/translation="MTQIGPDELEKAAVAVILKEPLSKRAEKAGDILVRLDL  
LGIRNPSAKITREIKSEGLPVLVNRVABGKBEKVDTEILVALSLKQSP  
DAVIDELASREDRDKVIAKAKHGVIISSHDFSTPSPOENTATLAMEFLAED  
AKIAPMGSMEDVNLKVLLEFKNIGTKVCTTIAMGRKHTRVAVLGSVLTYSI  
ESNAVAPGOLPDEVKIMEMLK"  
9843..10685  
/gene="arod"  
/note="synonym: MA4594"  
9843..10685  
/gene="arod"  
/codon\_start=1  
/transl\_table=1  
/product="shikimate 5-dehydrogenase"  
/protein\_id="AA07933.1"  
/db\_xref="GI:19918745"  
/translation="MKRVFGVDPVGHSLSPAMHNSAFALGMDCIYHAFVRPENTL  
RKAILGAEAMGFGGLNLTVPKBEALKLDFRPPLEBRIGAVNTVVSSEIGRGYN  
TDDIGARQALLBAVEIRGSKVVAVAGGAAVAPOLADGAEITVNRTEBAAVL  
AKDYAASLPGKINGTGLSKELIRPADLINTTIGMHPNTDTTITABELHSGIT  
VFQVYNPLETRLKEAKVAGAKTVSGVLMVYQGAAPFLMTGVEAPALMKTYLLE  
ALQA"  
10682..12112  
/gene="MA4595"  
10682..12112  
/gene="MA4595"  
/codon\_start=1  
/transl\_table=1  
/product="prephenate dehydrogenase"  
/protein\_id="AA07934.1"  
/db\_xref="GI:19918746"  
/translation="MNPDPKPSGKTKVLLGSGMGQFTFRFKERYKYTWCK  
GGKVEVARKLNPVASELDALIPENDIVIVSPINVTVEETIAEAPPMKAGSLMDT  
STRKVPVPAQAPAGVCEILGHPMGFPITIRGQTVLIPKESKEMFPIROL  
FEEGGAVEITTAEDRLVSVQGLTHFAVISGTTIDRLDFIKKSRKVPVSI  
MLDPVGRILGONPYLALIQENPGIPEVHAPEECEEELSLVRADEGFPVKMA  
AARKYDDTALARRSDKLINSRIAETVTLNLSGKCGFSHYVSGRVHGVKVGPN  
EVLITKLASGTSPIHINKFKIKLENTLRLSSELEMRERKLVHSVRDISVLIPG  
ADPEIVGAVKTDHVLVACQISDIYDGETEDKNEIRETRGLGVYRITTFQCNAD  
SVEAEVYALLCGLCRVERNKLTE"

## ORIGIN

Alignment Scores:  
Pred. No.: 0.00748 Length: 12180  
Score: 272.00 Matches: 76  
Percent Similarity: 45.63% Conservative: 44  
Best Local Similarity: 28.90% Mismatches: 123  
Query Match: 16.64% Indels: 20  
DB: 1 Gaps: 5

US-10-089-514-6 (1-322) x AE011181 (1-12180)  
QY 2 SerGlyPheProArgSerValValGlyGlySerGlyValValGlyGlyMetPheAla 21  
Db 10709 TCCGSAANAAGAGGTTTGTGATCTGGCGGAAGAGATGGGGCAGTGGTCTACT 10768  
QY 22 GlyLeuLeuArgGluAlaGlySerArgThrLeuVal-----ValAsp 35  
Db 10769 CGTTTTTTTAAAGAAAGGGGTATTAAGTTACGGTCTGGGAAAAGGGGCAAGTGAA 10828  
QY 36 LeuValProProProGlyArgProAlaCysLeuValGlyAspValThrAlaProGly 55  
Db 10829 GTGCCGAGAACTGATGATGATCCCTTTGCT----- 10858  
QY 56 ProGluLeuValAlaAlaLeuArgAspAlaAspLeuValLeuLeuValAlaHisGluAsp 75  
Db 10859 TCCGAACCTTGACCGACGATCCCTGAAAGCATATGATATGATATCCGATCCATCAC 10918  
QY 76 ValAlaLeuValAlaValAlaProValThrArgLeuMetArgProGlyAlaLeuLeuAla 95  
Db 10919 GTAACCGAAGACGACATTCGAGAAGTCGCCCGCAAGATGAACAGGAGACCTTCATG 10978  
QY 96 AspThrLeuSerValArgThrGlyMetValAlaGluLeuValAlaHisAlaPro---Gly 114  
Db 10979 GATTTTACCTCCACCAAGTAAAGCCTGTGAAGCCATGACAGAGTTTGACCTGCCGG 11038  
QY 115 ValGlnHisValGlyLeuAspPheAlaProAla---AlaGlyMetThrGlyArg 133  
Db 11039 GTTGAAGTATTCGGAAGCAGCCCATGTTGGGCCCGACATGCCACCATCAGGGGACAA 11098  
QY 134 ProVal-----AlaAlaValValThrArgAspGlyProGlyValThrAlaLeuLeuArg 151  
Db 11099 ACGGTATCTCTTGTCCCGTAAAGAGCGCTCGGAAAATATGTTCCGTAATCAGGACG 11158  
QY 152 LeuValGluGlyGlyGlyGlyArgProValArgLeuThrAlaGluGluHisAspArgThr 171  
Db 11159 CTTTGTGAGGAAGGCGGAGCCGACGCTTGAATATACAGACCGACCGAAGCAGACGCTG 11218  
QY 172 ThrAlaAlaThrGlnAlaLeuThrHisAlaValLeuLeuSerPheGlyLeuAlaLeuAla 191  
Db 11219 GTCTCGGTATACAGGGGCTGACGACATTTGCTTATATCTCAATAGGACGACCATTTGAC 11278  
QY 192 ArgLeuGlyValAspValAlaGlyAlaLeuAlaAlaThrAlaProProPheHisGlnValLeu 211  
Db 11279 AGCCTGATTTCCAGATCAAGAAATCCAGGAAATTCGTAAGTCCGCTCAGATATATG 11338  
QY 212 LeuAlaLeuLeuAlaArgValLeuGlyGlyGlySerProGluValThrGlyAspIleGlnArg 231  
Db 11339 CTGACTTTGTGGGCAAGATCTCTGGGCGAGACCCCTACTTACCGCCCTGATCCAGATG 11398  
QY 232 SerAspProArgAlaAlaAlaSerAlaArgArgAlaLeuAlaGluAlaLeuArgSerPheAla 251  
Db 11399 GAAACCCGGGATTCCTCCGAGGATGACATGAAGCATTTATAGAGAAATGTGAACCTTCC 11458  
QY 252 AlaLeuVal 254  
Db 11459 AGCCTGGTA 11467  
RESULT 14  
LOCUS AE013358  
DEFINITION Methanosarcina mazei strain Goel, section 140 of 379 of the  
complete genome.  
ACCESSION AE013358 AE008384  
VERSION AE013358.1 GI:20905730  
KEYWORDS  
SOURCE Methanosarcina mazei Goel  
ORGANISM Methanosarcina mazei Goel  
REFERENCE Methanosarcinaceae; Methanosarcinales;  
Methanosarcinaceae; Methanosarcina.  
1 (bases 1 to 9888)  
Depenmeier,U., Johann,A., Hartsch,T., Merkl,R., Schmitz,R.A.,



TITLE  
JOURNAL  
MEDLINE  
PUBMED  
22120827  
12125824  
2 (bases 1 to 9888)  
AUTHORS  
Deppe-meier U., Johann, A., Hartesch, T., Merkl, R., Schmitz, R. A.,  
Martinez-Arias, R., Heme, A., Wietzer, A., Baumer, S., Jacob, C.,  
Brieglebmann, H., Lienard, T., Christmann, A., Boemcke, M., Steckel, S.,  
Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H.-P.,  
Gunsalus, R. P., Fritz, H.-J. and Gottschalk, G.  
The genome of *Methanohalobium maza*: evidence for lateral gene  
transfer between bacteria and archaea  
J. Mol. Microbiol. Biotechnol. 4 (4), 453-461 (2002)  
TITLE  
JOURNAL  
Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute  
for Microbiology and Genetics, Grisebachstrasse 8, Goettingen  
37077, Germany  
FEATURES  
source  
1. .9888  
/organism="Methanohalobium maza Gol"  
/mol\_type="genomic DNA"  
/strain="Gol"  
/db\_xref="taxon:192952"  
242. .817  
/gene="MM1269"  
242. .817  
/gene="MM1269"  
/codon\_start=1  
/transl\_table=11  
/product="conserved protein"  
/protein\_id="AA030965.1"  
/db\_xref="GI:20905731"  
/translation="MKRRHATITISDKHWLKLTHAKPEPTOKALEIAECLENTTE  
QSELTVKQKYMCMCSIGSVCCQVKEALIMETVNHSEPKYVTKNPTETEIOIL  
LQKPLNCSLEEVNGLMTVFTSHMFDITDCKDNGDYLLILTHSLGINSKXILMT  
CKSLFTYTGADISKTSIGNIIFIKVFKDXTL"  
complement (1025. .1594)  
/gene="MM1270"  
complement (1025. .1594)  
/gene="MM1270"  
/EC\_number="6.3.5.2"  
/codon\_start=1  
/transl\_table=11  
/product="GMP synthase [glutamine-hydrolyzing]"  
/protein\_id="AA030966.1"  
/db\_xref="GI:20905732"  
/translation="MKRELKLVNNYNGQFCHLIRTVARDIMDKIIPNVTPIETILA  
EEDPGILSGSPMERAGLCFDYREIDVPIILICIGHQALALAGSHVAGKGGYA  
EIEVEVLIEDDILRGLGPKTTVASHADEVALIPDGFILHARSVDCEIEAMRHPTKPI  
YGVOMHPEVSHTKGSELLMNFVECDLY"  
2056. .2859  
/gene="MM1271"  
2056. .2859  
/gene="MM1271"  
/EC\_number="4.1.2.15"  
/codon\_start=1  
/transl\_table=11  
/product="2-dehydro-3-deoxyphosphoheptanote aldolase"  
/protein\_id="AA030967.1"  
/db\_xref="GI:20905733"  
/translation="MTMSEIGKKIRIERLMNRSSRMVITPLDHSIDGPIEGILNIT  
DTYRVAEAGANVLMOKGMYRGHSGYGHIDGLVYHISASVLSPPRAKVOCTVE  
EYIKMGADVASHINVSDETPADQLEMLGKISDCTCEMGPILAMNYPRGKKTINHD  
PVNVAAHARTGAEIGADVKTIVTGPDSRDVAGCPVPVVIAGSPKSTJDMLELM  
IDGAMGAGAGATIGRNIVFOHDPVRLTRAIETVHRRVBEALBOLK"  
3000. .4142  
/gene="MM1272"  
3000. .4142  
/gene="MM1272"  
/EC\_number="4.6.1.3"  
/codon\_start=1

/transl\_table=11  
/product="3-dehydroquinate synthase"  
/protein\_id="AA030968.1"  
/db\_xref="GI:20905734"  
/translation="MKKKSVMWKABEGGNEBQKERTTGLSEGACDVLNPNQVKNR  
ELGNITVAAPFARDNKSADIVVAGKRGEDGTRKPLPQELPSFVDNAATLLTDKGTIV  
GGYVVIKDRSYRFAAEMGKIDYLLVGTDMKVLPLENLIADLQREKVIIFGVSA  
EBARLAFQTLTEGADVLDSGNPOEIKDTIKARBLESESTELSAVTRVPELGMG  
DVCVDTCNLMORGEMLIGSOAGMFLVNSESDSPYVAPARFVNAGAVSHYKIG  
DKTRYLSELAGDSVTIIDSRGKREGIVGRIRIKIESRPLMLIEARAGDRTLLAIQNA  
ETIKLVKGGGTISVAKLKKGDEVLRLEBGRHKKKEITETIK"  
4203. .4931  
/gene="MM1273"  
4203. .4931  
/gene="MM1273"  
/EC\_number="4.2.1.10"  
/codon\_start=1  
/transl\_table=11  
/product="3-dehydroquinate dehydratase"  
/protein\_id="AA030969.1"  
/db\_xref="GI:20905735"  
/translation="MTQIGSFPLEKKTAVAVILKPKLENSKKAAGKADILERLD  
LGIRTPERAAYIRREIKARTGPIIVTSSGABGCKDCKEDRGLINLISLQDP  
DAIDIELSAGMKERKNKVIKAQDRETAIVSSHDPKTPLONMRTIIEEMPLAGADI  
AKLAVPLSVGDTLNLRLVTLDPKQAGSVCTIANGSQGKTRVVAPEYGSVLTASI  
ESNASAPQGLPVDKIKIMEMLK"  
4928. .5770  
/gene="MM1274"  
4928. .5770  
/gene="MM1274"  
/EC\_number="1.1.1.25"  
/codon\_start=1  
/transl\_table=11  
/product="Shikimate 5-dehydrogenase"  
/protein\_id="AA030970.1"  
/db\_xref="GI:20905736"  
/translation="MKQVGVGDPVGHSLSPAMNNAFSAIGMDCIYHAFVTPBEKL  
EKALIGAEAMGFGINLTVPLKEKALKDFVPDPLAKIGAVNTLVFEKGDIOGN  
TGPGKQKALDTPDVEIRGSRMVVAGGAARVAFOLAAGADITLVIRTEBRLVAG  
AREISAADLRKIRGTGSLKEILRDAVDVINTTIGMHPNTDATTIVAEELHSGLT  
VVDIYNPLETETLLREARTSGAKTISGVLMLVYQGAEARLMTGVPPELKKTYLLE  
ALQA"  
5767. .7185  
/gene="MM1275"  
5767. .7185  
/gene="MM1275"  
/EC\_number="1.3.1.12"  
/codon\_start=1  
/transl\_table=11  
/product="Prephenate dehydrogenase"  
/protein\_id="AA030971.1"  
/db\_xref="GI:20905737"  
/translation="MNTDYDAKTKVLLIGTGEMGQWTFTRFKERGVBYTWKGGK  
IEVAKKLDVPFALDLEAVIPESDVIIVSVINATEETIAELAPKKAGSILMDPFSIK  
VGVBEAMRKFAKQDVEIILGTHMPFGPTTPTINGQTVIIPVGVSEKMPVRLQFEE  
SGAHVEITTAEBDRILVSVGLTHFAVLAIGTIDRLDVKSRKFSVYSILND  
FVGRITLQGNPVLYALIQMENPGVPVHEAFICEKCELSLVAKHBEQFVRMKAAR  
KYDDTSHARRSDKILNSRIIEYETILNSTGVCGSRSHYSGSNHYGRLEKRPETV  
LMKLVSKTAPNPKRFTTLKLENLRPSEALREKRENLSHVRDLSVILPEGADP  
EVLRAVSTNKLADCKISDIYGVNGTLLERKGSYAEKGLTYRISIFGDDADDSVE  
TEVTALLCGLRIRKKNLKFS"  
7260. .8795  
/gene="MM1276"  
7260. .8795  
/gene="MM1276"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA030972.1"  
/db\_xref="GI:20905738"  
/translation="MSQVKKKISLLPALITTSIFPSLLISDKFTESKGNENVTAMA  
GVSGNOVSLMNRLETGSGEARLNAATELGNLGEPASALIEKMEENPGLGEIK  
SYMLLALLETGDKRAENIISYVLGKEASNAVADVQSGKEVPEDIMLAIOAKDKAM



RESLARSMTSEYNDITDLEALKEAEONSTLYTSFVLSNFSESENETILLOALKS  
EKGVYVSAAMALGOKKEKAAATPBLKMLADOPYLHAGSAVMALGEMDRANVTLMN  
ELKONKEDYIRRSSTALALGKLEAEAVPILERLDRASVRENSALVLCRMDETAV  
BPLKNTLESKDSERKOSLNTGVDRKSTVIALGIGTSTOTLIGVINDGEELP  
DVAVVAITIALGINSPEAVNLKTYENOMNNVNRGALIALGKTENQAAEFIK  
IGKDDGVSAAREALINNGETAVEPLIENLKTEDOKIKDETALILIBIGDRAVEPLIE  
AYK"

complement (8942. .9691)  
/gene="MM1277"  
complement (8942. .9691)  
/gene="MM1277"  
/EC number="2.1.1.-"  
/codon\_start=1  
/transl\_table=11  
/product="methyltransferase"  
/protein\_id="AA05073.1"  
/db\_xref="GI:20905739"  
/translation="MMETKEIIRNYMDHREISYTGIVESEERLAKMKLASKLGG  
RKLEILDVGTGQALAMFAMGHHLLTADVLSASMEKARKNALDINFIQDGA  
ENLOPDHFDVVSKEFLMLTDPKALLEMRVLLKKDMMIILADGWSGIFLKS  
IRTVSGIIRSVKERNRNDPFRKNTQULXNDLPLYSLKRGVFGLLDAGFEENIERM  
DSLCLSRKRGSLDKLDYSHPIYFIRAVK"

## ORIGIN

Alignment Scores:  
Pred. No.: 0.0301 Length: 9888  
Score: 256.00 Matches: 71  
Percent Similarity: 46.77% Conservative: 52  
Best Local Similarity: 27.00% Mismatches: 119  
Query Match: 15.66% Indels: 21  
DB: 1 Gaps: 5

US-10-089-514-6 (1-322) x AE013358 (1-9888)

QY 2 SerGlyPheProArgSerValValValGlySerGlyValValGlyGlyMetPheAla 21  
DB 5785 GCTGGAAAAACGAAGCTTCGATCTTGCGGGACCGAGAGATGGCGAGTGTGCACA 5844  
QY 22 GlyLeuLeuArgGlyValGlySerArgThrLeuVal-----ValAsp 35  
DB 5845 CGTTTTTTCAGAGAAAGAGCTTACGATCTCTGCGGGAAGCGCGCAAGATGAA 5904  
QY 36 LeuValProProProGlyArgProAspAlaCysLeuValGlyAspValThrAlaProGly 55  
DB 5905 GTGCGCAAAAACCTGACGCTACCTTTTGCTTTA----- 5937  
QY 56 ProGlyLeuValAlaAlaLeuArgAspAlaAspLeuValLeuValAlaHisGluAsp 75  
DB 5938 ---GACCTTAGGAGATTATCCCTGAGAAAGTACATATGATTCCGTCGATCAAT 5994  
QY 76 ValAlaLeuValAlaValAlaProValThrArgLeuMetArgProGlyAlaLeuLeuAla 95  
DB 5995 GCAACAGAGAAACCATCGAGAAATCGCCCAAAAATGAAAGCTGGAACCATCTTATG 6054  
QY 96 AspThrLeuSerValAlaGlyThrGlyMetAlaGlyLeuValAlaHisAlaPro---Gly 114  
DB 6055 GATTTTATCCCAATAAAGGTAGACCTGTGAACCGTAGAAAGTTTGCGCCCAAGAT 6114  
QY 115 ValGlnHisValGlyLeuAsnProMetPheAlaProAla---AlaGlyMetThrGlyArg 133  
DB 6115 GTTGAATTTCTTGGAGACGACCCCATGTTTGTCGACATTCACACATAAGGGGCA 6174  
QY 134 ProValAlaAlaValAlaThrArg-----AspGlyProGlyValThrAlaLeuLeuArg 151  
DB 6175 ACAATAATCCCTGTTCCGTAAGAGGTATTCGAAATAATGTTTCCGTAATCAGACAG 6234  
QY 152 LeuValGlnGlyGlyGlyArgProValArgLeuThrAlaGlnGlyHisAspArgThr 171  
DB 6235 CTTTTTAAAGAAAGCGCGGACGTTGAGATCAACAATCGAGGAAACGACGAGGCTG 6294  
QY 172 ThrAlaAlaThrGlnAlaLeuThrHisAlaValLeuLeuSerPheGlyLeuAlaLeuAla 191  
DB 6295 GTCTCACTAGTACGAGGGGCTTACTCATTTTGCTTATATTCGTATAGCCACACATTCAC 6354

QY 192 ArgLeuGlyValAlaAspValAlaArgAlaLeuAlaAlaThrAlaProProProHisGlnValLeu 211  
DB 6355 AGGCTGATTTTACGTCAGAAAAGTCACAGAAAATTCGTAGACCCGCTTTACACTATATATG 6414  
QY 212 LeuAlaLeuLeuAlaArgValLeuGlyGlySerProGlyValValGlyAspIleGlnArg 231  
DB 6415 CTCGACTTTGTGGCGAGATCTCCGGACAGAAATCCATCTCATGCGCCGATTCAGATG 6474  
QY 232 SerAsnProArgAlaAlaAspAlaArgAlaAlaLeuAlaGlyAlaLeuArgSerPheAla 251  
DB 6475 GAAACCCCTGAGAGTCTCTGAGTGCATGAAGCATTTATAAAGAGTGTGAGAGCTTTCC 6534  
QY 251 AlaAlaLeu 253  
DB 6535 AGCCTTG 6541

RESULT 15  
BX957223  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE

JOURNAL

FUNDED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

gene

CDS

1.268122  
/organism="Methanococcus maripaludis S2"  
/mol\_type="genomic DNA"  
/strain="S2"  
/db\_xref="taxon:267377"  
227..1555  
/gene="secY"  
/locus\_tag="MMP1422"  
227..1555  
/gene="secY"  
/locus\_tag="MMP1422"  
227..1555  
/locus\_tag="MMP1422"  
/evidence=not experimental  
/transl\_table=11  
/product="Aldehyde dehydrogenase:SecY protein"  
/protein\_id="CAF30978.1"  
/db\_xref="GI:45047860"  
/db\_xref="UniProt/TREMBL:O6LKD0"  
/translation="MESFLIKIKPLLELIPYKRPRLRISPKKLOMTALVTVLPIL  
GTTDITVGSSENPATFDPWQTYTASMGTLITIGCPITYTAAGINOLVGSBLISDL  
SKPMNALRQGLQKLFGLALCPLEALMFVGAAGFALVTLMTAVLVFQALGAILITY  
LDEIVSRVIGSGIGLFIAGVQSITFVGTFGABEGVLMKFTFAMRAGSLMTALVILP  
ILGTTIVLVVVVVEISIRVEIPLAHGRVGAAGVGIKPIYVSNLPLVLAALFAMIO  
LMGMFLKKGFPLIGHYTGRAVDGLAAYFSTPVGISSTADPLAHAVFVTVMMVVICI  
LRELFWERTSGIDAKSMARKKGLNDMAIINGPRKSQISIQRLKRYITPITVMSARVG



misc\_feature  
FLAAADFTGALGGTGVLLTVTSIVRYLVEQLVQQLSEIHPISAKPIRK"  
order(317, .385,443, .511,572, .640,650, .709,728, .796,  
854, .922,995, .1063,1181, .1249,1352, .1420,1430, .1489)  
/gene="secy"  
/locus\_tag="MMP1422"  
/note="10 probable transmembrane helices predicted by  
TMHMM2.0"  
complement(1595, .2494)  
/locus\_tag="MMP1423"  
complement(1595, .2494)  
/locus\_tag="MMP1423"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="putative glycosyl transferase, family 4:Aldehyde  
dehydrogenase"  
/protein\_id="CAF30979.1"  
/db\_xref="GI:45047861"  
/db\_xref="UniProt/TREMBL:O6LXC9"  
/translation="MDLYEVLALFCFVLSLFTKEMIKKVVYKYGYDLHKADKIKV  
AEWGLSPVYVSVMLFENPALSIFLPGFVGVIDISRLSKKIKVLTPLGPPV  
APFLKLGFLPSILLIIGIFVSSNLTMNLGPNGLIGMGLICLMAVVCIMMGIDFG  
FKLLIFLAAYLGLIYNNRYPAKVPFGDTGTEPLGALFLATIAVRGPITPLMIFP  
MDALFKQFTAGVTKDSVFTPTQLKNGKLCVGGYLSFPMILMKKAMEYKIVLV  
MALEAFGLISILYTRFENIP"  
complement(order(1613, .1678,1853, .1918,1964, .2020,  
2042, .2107,2153, .2218,2255, .2320,2423, .2479))  
/locus\_tag="MMP1423"  
/note="7 probable transmembrane helices predicted by  
TMHMM2.0"  
2591, .5149  
/locus\_tag="MMP1424"  
2591, .5149  
/locus\_tag="MMP1424"  
/BC\_number="2...-"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="Oligosaccharyl transferase, SRT3 subunit"  
/protein\_id="CAF30980.1"  
/db\_xref="GI:45047862"  
/db\_xref="UniProt/TREMBL:O6LXC8"  
/translation="MGEFLNKVSDPFKKKEKIKIILLFTIGMSFOIRAOQDMAFT  
DNSYLODMFDDNMYLTLADPYLLRMTEYNNVDYSVGETTIGDENGIPYTI  
QVAPGREAGLVALLSLATVLYVSWNSIDSTVTIMAAFWAPAIMSIFLGPVEFIV  
RRNTASNGIGLVALLISPSLLYKTSAGSPDTPLEILLPLFTVMMEALHGBEN  
SKKSGIFGGIAAILIGIYPMMSGMWYAPITGPIVLTYAYEILKSKLKVITTS  
LITLVGAILVLSLSTGISCFINMILSPIGTVINEAKITGMNVNMYVSELAIPVT  
DIENSVGNITMLISGILSLSPFPGDKQKIDIKVALYTLMLIATVYATKQIR  
FVALMTPALAIGIGIAGQIENIKREKKEVEILYVIGILSVITLIKXGGLFNIL  
VPTTVPPIAVYISIIALLVLAAYKIIDITSEKQAVKAVGILLAFMLPPSMAAIVP  
FYTAPTMNNGMDSLSMISKEPSPENSVCMMNGHITWMAATPKMTTPGQSNTPRA  
YWGAFSTSDENLSVGLIRMLATSGDSAYDDSIITKKTGSIKDTVIDINKLPLR  
TEAKASLVNNYDLTDAEAEVLDLTHPKVTNPBYLLTYNNMTSIAVSWMFGWNPSL  
PASTENSDEMGGYQOLGSAODINGTIVVYIQLQBTDSRVNINLEIITSELSKANA  
VIDSNQTSQMSQPNFKLILKVNQNVYEQETNENGDSYSEIVLEKLSDEGYVYAAVS  
SKULEDSIVTKLHFLDYGLKESISEKESVDPRSYGIQPGFKVYSVDYGRDYN"  
2591, .2701  
/locus\_tag="MMP1424"  
/note="Signal predicted by SignalP 2.0 HMM (Signal peptide  
probability 0.992) with cleavage site probability 0.966 at  
residue 37"  
order(2648, .2707,2999, .3067,3086, .3145,3173, .3232,  
3257, .3310,3320, .3379,3413, .3481,3611, .3670,3704, .3754,  
3767, .3826,3863, .3922,3950, .4018,4055, .4123)  
/locus\_tag="MMP1424"  
/note="13 probable transmembrane helices predicted by  
TMHMM2.0"  
5262, .5801  
/locus\_tag="MMP1425"  
5262, .5801  
/locus\_tag="MMP1425"  
/codon\_start=1

/evidence=not\_experimental  
/transl\_table=11  
/product="Protein of unknown function DUF127"  
/protein\_id="CAF30981.1"  
/db\_xref="GI:45047863"  
/db\_xref="UniProt/TREMBL:O6LXC7"  
/translation="MAIEILRLGHRGEBDRKISTHVALTSRALGAEKIIFTEEDKHVK  
ESVERIIVEMAGEGFEVYKSWRSYAKPKDNGIIVHLLTMGENINKIMTEIRIDISK  
TKNKLILLIIGAERVPREAYDLADYNLSVGNQHSVLAALAIRLDPFTECKTLTSEYSD  
AKIKVTPSKSEKCVPEKD"  
5825, .6445  
/gene="dcd"  
/locus\_tag="MMP1426"  
/BC\_number="3,5,4,13"  
/note="Citation: J. Biol. Chem. 278:11100-11106(2003)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="Bifunctional dCTP deaminase/dUTP diphosphatase"  
/protein\_id="CAF30982.1"  
/db\_xref="GI:45047864"  
/db\_xref="UniProt/TREMBL:O6LXC6"  
/translation="MILSDKDIENVYTSKRVLIPEPNSKYGPCSYDVTIGSEFIKYN  
EDVYDVKKTINSKRTIENSVMVCCPLNHLDDTTIKNTKEKNVDCVSGSLGTTNE  
YVELPNDVCAQYQGRSSFGRVFLQTHQAGWIDSGFKITLIEIAYADPVLVYKNQR  
VQQLIFSXTLSPADIGYSDRCKSKYAGQTSVWASDIKDFETNEBE"  
6449, .7825  
/locus\_tag="MMP1427"  
6449, .7825  
/locus\_tag="MMP1427"  
/BC\_number="3,5,4,13"  
/note="Citation: J. Biol. Chem. 278:11100-11106(2003)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="Conserved Hypothetical Protein"  
/protein\_id="CAF30983.1"  
/db\_xref="GI:45047865"  
/db\_xref="UniProt/TREMBL:O6LXC5"  
/translation="MFVPEEIIETIKMTEYQMLDIRTTTGINLKDCAKDLDLKEN  
YDKTSLCGNLVETNAKESOKGYPIVNRKISTPIGLIMSTYKGLSDEBAVACV  
EKGITLDAKRVGDFGTGVSALVOKRATYBEKMLISIPMLTKTDKVCASVAVAT  
TKAGINMYVKRMGSEIVETSETITDAIGCATIYVFCNAPENPNMAAGFHPGSGDA  
VTNAGVSGPVRAVVEQLKQKQIGVSDIEIKTIAFKITRMGELVGEVANSLVGVFG  
IVDLSLAPTPAIGDSIANTLEAVGLERCGTHGTALMLNDVAKKGMASSNVGL  
SGAFLPVSEDAQMIIEAVEGALRLKLEAMTCVSGIDMLIVPEKTPASTLSAIMAD  
EMAGIMINKTTAAVNIIPVPGKDVGDYVEYGLLGTAPIMPVSEFSSBELIBRGRIIP  
APIGSLIN"  
7898, .8059  
/locus\_tag="MMP1428"  
7898, .8059  
/locus\_tag="MMP1428"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="Hypothetical protein"  
/protein\_id="CAF30984.1"  
/db\_xref="GI:45047866"  
/db\_xref="UniProt/TREMBL:O6LXC4"  
/translation="MAKNQDAGLSTAGLVRYMDDVSKIKIAPKVLGLTVSIIIE  
ATLVNIGILIT"  
7994, .8053  
/locus\_tag="MMP1428"  
/note="1 probable transmembrane helix predicted by  
TMHMM2.0"  
complement(8114, .8431)  
/gene="rpm"  
/locus\_tag="MMP1429"  
complement(8114, .8431)  
/gene="rpm"



## Alignment Scores:

Pred. No.:	3.24	Length:	268122
Score:	239.00	Matches:	67
Percent Similarity:	47.47%	Conservative:	55
Best Local Similarity:	26.07%	Mismatches:	127
Query Match:	14.62%	Indels:	8
DB:	1	Gaps:	5

US-10-089-514-6 (1-322) x BK957223 (1-268122)

```

QY      9 ValValGlyGlySerGlyAlaValGlyGlyMetPheAlaGlyLeuLeuArgGluAlaGly 28
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80133 ATATCGCGCGAACTGACGGGCTTGGAATAATGTTGCATCATTTCTTAATAAATAGGGGT 80192

QY      29 SerArgThrLeuValValaLeuLeuValProProProGlyArgProAspAlaCysLeuVal 48
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80193 TACGATGTACTCGTATCTGGAGAGGATGTATTAAGAAAAGAAAGTTGAAAAGAACTT 80252

QY      49 GlyAspValThrAlaProGlyProGlyLeuAlaAlaLeuArgAspAlaAspLeuVal 68
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80253 GGT---GTAACTTACACAAATAACAAATATGATGCTGCA---GAAAAAGAGATATTGTA 80306

QY      69 LeuLeuAlaValHisGluAspValAlaLeuLysAlaValAlaProValThrArgLeuMet 88
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80307 ATATTTCTGTCCCTTAATAATGTACAGAACTGTGATTAAGAAAGTTGCTCCGCATGTT 80366

QY      89 ArgProGlyValaLeuLeuAlaAspThrLeuSerValArgThrGlyMetAlaAlaGluLeu 108
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80367 AAAAAAGTTCACTTTTAAATGACATTAACATCGATTAAGAAAAGAAAGCCCTGCAAAACGTGATG 80426

QY      109 AlaAlaHisAla---ProGlyValaGluHisValaGlyLeuAsnProMetPheAlaProAla 127
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80427 GCAGAAATTTCAAAAGAACTGAGATTTGTACTTCCGACACACCTATGTTGGGCCAGAA 80486

QY      128 AlaGlyMetThrGlyArgProValAlaAlaValaVala-----ThrArgAspGly 143
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80487 ACTCCTTCGTTAAACCGGAGGTTGTAACTTAACACTTAACCTTAAGAAAAGAAAACAAT 80546

QY      144 ProGlyValaThrAlaLeuLeuArgLeuValGluGlyGlyGlyArgProValArgLeu 163
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80547 CCATTTTGGAAAAAATTAAGAATTTTGAAGAAATGAAAGTGCAAAAGTAATTTGTTGTA 80606

QY      164 ThrAlaGluGluHisAspArgThrThrAlaAlaThrGluAlaLeuThrHisAlaValaLeu 183
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80607 TCCCAAAAGAACACGATTAATTAATGCGTGTAGTTCAAGGCTTAACCTCATTTTGTATAT 80666

QY      184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValaAspValArgAlaLeuAlaAlaThr 203
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80667 ATTTCACCTTGATCCCACTTAAGAAATCTTGGAATAGATATAAAGAAATCCAGAACTTT 80726

QY      204 AlaProProProHisGluValaLeuLeuAlaLeuLeuAlaArgValaLeuGlyGlySerPro 223
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80727 GCCTCCCAATTTACGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 80786

QY      224 GluValaTyrglyAspIleGluArgSerAsnProArgAlaAlaSerAlaArgAlaLeu 243
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80787 TACCTATATGACATATCCAGATGACACACCCCAATTAATAAACATTTCAGAAACATTC 80846

QY      244 AlaGluAlaLeuArgSerPheAlaAlaLeuValaGlyAspAspProAspArg 260
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80847 ATTAATAAACTGCAATAATTAATTAAGTAATTC---CAGATAAGCATAGG 80894

```

Search completed: October 5, 2005, 15:36:38  
 Job time : 5131.02 secs



This Page Blank (uspio)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 4, 2005, 21:38:23 ; Search time 565.456 Seconds  
(without alignments)  
3371.006 Million cell updates/sec

Title: US-10-089-514-6  
Perfect score: 1635  
Sequence: 1 MSGFRSVYVGGSGAVGGMF.....HCOELFRTIHRDDEGKDR 322

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues  
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_p2n.model -DEV=xlh  
-O/cgn2.1/USPTO.spool.h/US10089514/runat.04102005.105743.8027/app.query.fasta\_1.1621  
-DB=N Geneseg -QFMT=fastcap -SUFFIX=ring -MINMATCH=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdt -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcio -NORM=ext -HEAFSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10089514 @CGN 1.1 860 @runat.04102005.105743.8027 -NCPUs=6 -ICPU=3  
-NO WMAP -LARGEOUTERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
N\_Geneseg\_16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1635	100.0	969	4	AAAF6411 4-amino-4
2	1635	100.0	969	8	AAAL50182 S venezue
3	1588	97.1	3305	8	ABZ69799 Plasmid p
4	1588	97.1	12391	8	ABZ69798 Plasmid p
5	661.5	40.5	888	2	AAT59269 Streptomy

C	6	661.5	40.5	2888	2	AAT59268	AAT59268 Streptomy
	7	382	23.4	888	10	ACF71774	ACF71774 Photorhab
	8	382	23.4	110000	10	ACF67367_49	Continuation (50 o
	9	382	23.4	110000	10	ACF65387_1	Continuation (2 of
	10	333	19.8	110000	11	ADM27081_07	Continuation (8 of
	11	237	14.5	1122	8	AAD50753	Aad50753 Escherich
	12	235	14.4	110000	2	AAV21209_09	Continuation (10 o
	13	233	14.3	1122	6	ABN83448	ABn83448 Escherich
	14	233	14.3	1238	6	ABA99612	ABA99612 E. coli t
	15	233	14.3	1238	6	AAI46032	AAI46032 E. coli t
	16	227	13.9	1131	11	ACH98777	ACH98777 Klebsiell
	17	225.5	13.8	1122	8	AAD50752	Aad50752 Escherich
	18	225.5	13.8	1122	12	ADJ98178	Adj98178 Pantoea a
	19	225.5	13.8	1152	6	ABN83449	ABn83449 Escherich
	20	224	13.7	3957	13	ADT05414	Adt05414 Haemophil
	21	224	13.7	349980	13	ADT05737	Adt05737 Haemophil
	22	216	13.2	110000	2	AAT42063_13	Continuation (14 o
	23	213	13.0	1792	5	AA65801	AA65801 DNA encod
	24	200	12.2	1125	10	ADP00853	Adp00853 Bacterial
	25	177.5	10.9	1122	10	ACF69143	ACf69143 Photorhab
	26	177.5	10.9	110000	10	ACF67367_20	Continuation (21 o
	27	177.5	10.9	243072	10	ACF65382	ACf65382 Photorhab
	28	165	10.1	2025	13	AD661947	Ad661947 Bacterial
	29	163.5	10.0	110000	4	AAI99682_39	Continuation (40 o
	30	159	9.7	88421	6	AAI40781	AAI40781 Bacteri
	31	152	9.3	891	13	ADR31472	Adr31472 Choriomat
	32	150.5	9.2	1362	8	ABX56041	ABx56041 M. echino
	33	150.5	9.2	36602	13	ADR01210	Adr01210 Farnesyl
	34	150	9.2	42291	9	ACC58253	ACC58253 Chlorobac
	35	149	9.1	1236	9	ADA37164	Ada37164 Human SUV
	36	149	9.1	2732	2	AAT43625	Aat43625 Chromatin
	37	149	9.1	2733	12	ADP10568	Adp10568 Reference
	38	149	9.1	71138	12	AD013830	Ado13830 dbv gene
	39	148	9.1	1023	13	ADR41634	Adr41634 Lipid acy
	40	148	9.1	1023	13	ADR41584	Adr41584 Lipid acy
	41	147.5	9.0	12152	5	AA508699	AA508699 M. carbon
	42	147.5	9.0	37116	10	ABZ66810	ABz66810 Orthosomy
	43	147	9.0	73882	13	ABD07351	ABd07351 tcp gene
	44	146.5	9.0	2376	11	ABD08026	ABd08026 Pseudomon
	45	146	8.9	2120	10	ADG91050	Adg91050 Hepatic s

## ALIGNMENTS

RESULT 1	AAAF6411	standard; DNA; 969 BP.
ID	AAAF6411	
XX	AAAF6411	
AC	AAAF6411	
XX		
DT	22-JUN-2001	(first entry)
XX		
DE	4-amino-4-deoxyphenolic acid dehydrogenase coding sequence.	
XX		
KW	Metabolite; benzene; chorismic acid; p-aminophenylpyruvic acid;	
XX		
KW	4-amino-4-deoxyphenolic acid dehydrogenase; enzyme; papC; ds.	
XX		
OS	Streptomyces venezuelae.	
XX		
PH	Key	Location/Qualifiers
FT	CDS	1..969
FT		/tag= a
XX		/product= "4-amino-4-deoxyphenolic acid dehydrogenase"
XX		
PN	MO200123542-A1.	
XX		
PD	05-APR-2001.	
XX		
PF	29-SEP-2000; 2000WO-JP006783.	
XX		
PR	29-SEP-1999; 99JP-00276314.	
XX		
PA	(MEIJ ) MEIJI SEIKA KAISHA LTD.	



XX Yanai K, Okakura K, Yasuda S, Watanabe M, Miyamoto K, Mido N;  
 PI Murakami T;  
 XX  
 XX WPI; 2001-290517/30.  
 DR P-PSDB; AAB82073.  
 PT Transformant producing secondary metabolite modified with functional  
 PT group e.g. benzene with nitrogen-containing substituent at para-position,  
 PT PF1022, with ease at low cost, for application in pharmaceuticals and  
 PT agrochemicals.  
 XX  
 PS Claim 15; Page 67-70; 83pp; Japanese.  
 XX  
 XX The present invention relates to a transformant having been modified so  
 CC as to produce a secondary metabolite. The secondary metabolite has a  
 CC benzene ring skeleton free from substitution at the para-position by a  
 CC nitrogen-containing functional group, thereby enabling the production of  
 CC a secondary metabolite with a benzene ring skeleton substituted at the  
 CC para-position by a nitrogen-containing group. The transformant organism  
 CC of the present invention has been produced by transferring a gene  
 CC participating in the biosynthesis pathway from chorismic acid into p-  
 CC aminophenylpyruvic acid. The present sequence is the coding sequence for  
 CC 4-amino-4-deoxyphenolic acid dehydrogenase (pdc), from Streptomyces  
 CC venezuelae. pdc participates in the biosynthesis pathway from chorismic  
 CC acid into p-aminophenylpyruvic acid, and so the pdc gene can be used to  
 CC produce the transformant of the present invention. The transformant can  
 CC be used to produce metabolites for application in pharmaceuticals,  
 CC veterinary drugs and agrochemicals  
 XX  
 SQ Sequence 969 BP; 106 A; 397 C; 358 G; 108 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 8.02e-88 Length: 969  
 Score: 1635.00 Matches: 322  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 4 Gaps: 0

US-10-089-514-6 (1-322) x AAF86411 (1-969)

QY 1 MetSerGlyPheProArgSerValValValGlySerGlyAlaValGlyMetPhe 20  
 DB 1 ATGAGCGGCTTCCCGGAGCGCTGCTGCGGCGAGCGGGCGGCGGCATGTC 60  
 QY 21 AlaGlyLeuLeuArgGluAlaGlySerArgThrLeuValValAspLeuValProPro 40  
 DB 61 GCGGGCTGCTGGGAGGGGCGAGCGGACCGCTGCTGCACTCGTACCGCGCG 120  
 QY 41 GlyArgProAspAlaCysLeuValGlyAspValThrAlaProGlyProGluLeuAla 60  
 DB 121 GAGCGCGCGGACCGCTGCTGCGGAGCGTCAACCGCGCGCGCGCACTCGGGCC 180  
 QY 61 AlaLeuArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuVal 80  
 DB 181 GCCCTCCGGGACCGGACCTCGCTGCTGCGCGTACAGAGACCTGGCGCTCAAGGCC 240  
 QY 81 ValAlaProValThrArgLeuMetArgProGlyValAlaLeuLeuAlaAspThrLeuSerVal 100  
 DB 241 GTGGCGCGGTGACCGGCTCATGCGGCGCGGCGCTGCTCGACACCGCTGTCGTC 300  
 QY 101 ArgThrGlyMetAlaAlaGluLeuValAlaHisAlaProGlyValGlnHisValGlyLeu 120  
 DB 301 CGGACGGGCGATGCGCGGAGCTCGGCGCCACCGCGCTCGACACGTCGGGCTTC 360  
 QY 121 AsnProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValThr 140  
 DB 361 AACCCGATGTTCCCGCGCGCGCGGCAATGACCGGCGACCGCTGCGCGCGTGCACC 420  
 QY 141 ArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGluGlyGlyGlyValGlyPro 160  
 DB 421 AGGAGCGGCGCGGCTGCAAGCGCGCTGCGGCTGCTCGAAGGCGGCGCGCGCGCC 480

QY 161 ValArgLeuThrAlaGluGluHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHis 180  
 DB 481 GTACGGCTCAGCGCGGAGGACGACGCGGACGAGCGGCGGCCACCGACCTGACGCAC 540  
 QY 181 AlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeu 200  
 DB 541 GCGGTGCTCTCTCTCTCGGGCTCGCGCTCGCGCTCGCGCTCGGCTCGACCTCGGGCTTG 600  
 QY 201 AlaAlaThrAlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGly 220  
 DB 601 GCGGCGCGGACCGCGCGCCCAACGAGTCTGCTGCGCTCTGCGCTGCTGCTGCTGCGC 660  
 QY 221 GlySerProGluValTyrGlyAspIleGlnArgSerAsnProArgAlaAlaSerAlaArg 240  
 DB 661 GCGAGCGCGGAGGTGACGGGAGCATCCAGCGGTCCAAACCGCGGCGGCGTCCGCGCGC 720  
 QY 241 ArgAlaLeuAlaGluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArg 260  
 DB 721 CGGGCGCTTGGCGAGCGCTTGGCTTGGCGGCTGCTGCGGCGAGCACCGGACCGT 780  
 QY 261 AlaAspAlaProGlyArgAlaAspAlaProGlyHisProGlyGlyCysAspGlyAlaGly 280  
 DB 781 GCGAGCG 840  
 QY 281 AsnLeuAspGlyValPheGlyGluLeuArgArgLeuMetGlyProGluLeuAlaGly 300  
 DB 841 AACCTGACGAGGCTCTTGGGGAACCTCGCGCTCATGAGACCGGAGCTCGCGCGCGCG 900  
 QY 301 GlnAspHisCysGlnGluLeuPheArgThrLeuHisAlaArgThrAspAspGluGlyVal 320  
 DB 901 CAGGACCACTGCGCAGAGAGCTGTTCGACACCTTCACCGCACCGAGCGAGCGAGAG 960  
 QY 321 AspArg 322  
 DB 961 GACCGA 966

## RESULT 2

AAL50182  
 ID AAL50182 standard; DNA; 969 BP.

AC AAL50182;

DT 28-JAN-2003 (first entry)

DE S venezuelae PF1022 substance gene #3.

KM Substance PF1022; biosynthesis; transgenic; chorismic acid; anthelmintic;

KW phenylalanine; p-aminophenylpyruvic acid; pharmaceutical;

KX veterinary drug; gene; ds.

OS Streptomyces venezuelae.

FX Key Location/Qualifiers

FT CDS 1..969 /tag= a

FT /product= "PF1022 substance"

XX MO20027244-A1.

XX 03-OCT-2002.

XX 22-MAR-2002; 2002WO-JP002782.

XX 22-MAR-2001; 2001JP-00082227.

XX (MEIU ) MEIJI SEIKA KAISHA LTD.

XX Yanai K, Sumida N, Watanabe M, Moriya T, Murakami T;

DR WPI; 2003-018934/01.

DR P-PSDB; AAO19565.

XX



PT Novel biosynthesis gene-transferred transformants for producing PF1022  
 PT substance derivatives by fermentation, as pharmaceuticals or veterinary  
 PT drugs with antihelmintic activity.

XX Claim 17, Page 87-90; 116pp; Japanese.

XX The present invention relates to transformants capable of producing  
 CC PF1022 substance derivatives. These were obtained by transferring a gene  
 CC participating in the biosynthesis pathway from chorismic acid to p-  
 CC aminophenylpyruvic acid (biosynthesis gene) into a phenylalanine-  
 CC requiring host derived from an organism producing the PF1022 substance.  
 CC The transformants are producing PF1022 substance derivatives by  
 CC fermentation, for use as pharmaceuticals or veterinary drugs. The present  
 CC sequence is a substance PF1022 coding sequence from Streptomyces  
 CC venezuelae

XX Sequence 969 BP; 106 A; 397 C; 358 G; 108 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.:	8,02e-88	Length:	969
Score:	1635.00	Matches:	322
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-10-089-514-6 (1-322) x AAL50182 (1-969)

ID	Sequence	Score	Length	Matches	Conservative	Mismatches	Indels	Gaps
QY	1 MetSerGlyPheProAgservAlaValGlyGlySerGlyAlaValGlyGlyMetPhe	1635.00	969	322	0	0	0	0
DB	1 ATAGAGCGGCTTCCCGGACGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCTTTC	100.00%	969	322	0	0	0	0
QY	21 AlaGlyLeuLeuArgGlyAlaGlySerArgThrLeuValAlaAspLeuValPropPro	100.00%	969	322	0	0	0	0
DB	61 GCCGGGCTGCTGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	100.00%	969	322	0	0	0	0
QY	41 GlyArgProAspAlaCysLeuValGlyAspValThrAlaProGlyProGlyLeuAla	100.00%	969	322	0	0	0	0
DB	121 GAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	100.00%	969	322	0	0	0	0
QY	61 AlaLeuArgAspAlaAspLeuValLeuLeuAlaValHisGlyAspValAlaLeuVal	100.00%	969	322	0	0	0	0
DB	181 GCCCTCCGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	100.00%	969	322	0	0	0	0
QY	81 ValAlaProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSer	100.00%	969	322	0	0	0	0
DB	241 GTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	100.00%	969	322	0	0	0	0
QY	101 ArgThrGlyMetAlaAlaGlyLeuAlaAlaHisAlaProGlyValGlnHisValGly	100.00%	969	322	0	0	0	0
DB	301 CGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	100.00%	969	322	0	0	0	0
QY	121 AspProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValThr	100.00%	969	322	0	0	0	0
DB	361 AACCCGATGTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	100.00%	969	322	0	0	0	0
QY	141 ArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGlyGlyValArgPro	100.00%	969	322	0	0	0	0
DB	421 AGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	100.00%	969	322	0	0	0	0
QY	161 ValArgLeuThrAlaGlyGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThr	100.00%	969	322	0	0	0	0
DB	481 GTTCCGCTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	100.00%	969	322	0	0	0	0
QY	181 AlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuValArgValAlaGly	100.00%	969	322	0	0	0	0
DB	541 GCCGTGCTCTCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	100.00%	969	322	0	0	0	0
QY	201 AlaAlaThrAlaProProProHisGlnValLeuLeuAlaLeuLeuAlaGlyVal	100.00%	969	322	0	0	0	0
DB	601 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	100.00%	969	322	0	0	0	0
QY	221 GlySerProGlyValTyrglyAspIleGlnArgSerAspProArgAlaSerAlaArg	100.00%	969	322	0	0	0	0

ID	Sequence	Score	Length	Matches	Conservative	Mismatches	Indels	Gaps
DB	661 GGCAGCGCCGAGGTGATACGGGACATCCAGCGGTCCACCCCGGCGGCGGCGGCGG	100.00%	969	322	0	0	0	0
QY	241 ArgAlaLeuAlaGlyAlaLeuArgSerPheAlaAlaLeuValGlyAspProAspArg	100.00%	969	322	0	0	0	0
DB	721 CGGCGGCTGCTCCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	100.00%	969	322	0	0	0	0
QY	261 AlaAspAlaProGlyValThrAlaProGlyValHisProGlyGlyCysAspGlyAlaGly	100.00%	969	322	0	0	0	0
DB	781 GCCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	100.00%	969	322	0	0	0	0
QY	281 AspLeuAspGlyValPheGlyGlyLeuLeuArgLeuMetGlyProGlyLeuAlaGly	100.00%	969	322	0	0	0	0
DB	841 AACCTGAGCGGCTTCCCGGGAATCCGCGGCTCATGGGACCGAGGCTCGCGGCGG	100.00%	969	322	0	0	0	0
QY	301 GlnAspHisCysGlnGlyLeuPheArgThrLeuHisArgThrAspArgGlyGly	100.00%	969	322	0	0	0	0
DB	901 CAGAGCCACTGCCAGGAGGTGTTCCGACCTTCACCGGACCGACGAGAGCGAGAG	100.00%	969	322	0	0	0	0
QY	321 AspArg 322	100.00%	969	322	0	0	0	0
DB	961 GACCGA 966	100.00%	969	322	0	0	0	0

RESULT 3  
 ABZ69799  
 ID ABZ69799 standard; DNA; 3305 BP.  
 XX  
 AC ABZ69799;  
 XX  
 DT 08-APR-2003 (first entry)  
 XX  
 DE Plasmid papABC.  
 XX  
 KW DHFR; translation; orthogonal tRNA; O-tRNA; O-RS; CAT; TyRS;  
 KW orthogonal aminoacyl tRNA synthetase; unnatural amino acid;  
 KW chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;  
 KW biosynthesis; p-aminophenylalanine; pAF; chorismate.  
 OS Synthetic.  
 OS  
 XX  
 PN WO200285923-A2.  
 XX  
 PD 31-OCT-2002.  
 XX  
 PF 19-APR-2002; 2002WO-US012465.  
 XX  
 PR 19-APR-2001; 2001US-0285030P.  
 PR 06-FEB-2002; 2002US-0355514P.  
 XX  
 PA (Scrip) SCRIPPS RES INST.  
 XX  
 XX Schultz P, Wang L, Anderson JC, Chin JWK, Magliery TJ;  
 PI Meggers EL, Mehl RA, Pasternak M, Santoro SW, Zhang Z;  
 XX  
 DR WPI; 2003-120430/11.  
 XX  
 PT Composition useful for producing protein comprising unnatural amino acid.  
 PT has translation system comprising orthogonal tRNA and orthogonal  
 PT aminoacyl tRNA synthetase.  
 XX  
 XX Example 4; Page 127-128; 188pp; English.  
 XX  
 CC The invention relates to a novel composition comprising a translation  
 CC system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl  
 CC tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O  
 CC -tRNA with at least one unnatural amino acid in the translation system  
 CC and the O-tRNA recognises at least one selector codon. A composition of  
 CC the invention is useful for producing at least one protein comprising at  
 CC least one unnatural amino acid. The protein is the Asp121G mutant of  
 CC chloramphenicol acetyltransferase (CAT), the Tyr163Tg mutant of mouse  
 CC dihydrofolate reductase (DHFR), or a Tyr163Tg mutant of mouse  
 CC dihydrofolate reductase comprising CHOON His6tag. The unnatural amino  
 CC acid is provided exogenously. The translation system is a cell and the



CC unnatural amino acid is biosynthesised by the cell. The present sequence  
CC represents a plasmid containing the individual genes papABC that encode  
CC the enzymes used to carry out the conversion of chorismate to the  
CC unnatural amino acid p-aminophenylalanine (pAF)

XX Sequence 3305 BP; 445 A; 1327 C; 1121 G; 412 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	1,49e-84	Length:	3305
Score:	1588.00	Matches:	315
Percent Similarity:	96.65%	Conservative:	2
Best Local Similarity:	96.04%	Mismatches:	5
Query Match:	97.13%	Indels:	6
DB:	8	Gaps:	1

US-10-089-514-6 (1-322) x ABZ69799 (1-3305)

```
QY      1 MetSerGlyPheProAArgSerValValGlyGlySerGlyAlaValGlyGlyMetPhe 20
        |||||
DB      2007 ATGAGCGGCTTCCCGGAGCGCTGCTGCTGCGGAGCGGAGCGGTGGCGCATGTTTC 2066

QY      21 AlaGlyLeuLeuArgGlyAlaGlySerArgThrLeuValAlaSerLeuValProProPro 40
        |||||
DB      2067 GCCGGGCTGCTGCGGAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 2126

QY      41 GlyArgProAspAlaCysLeuValGlyAspValThrAlaProGlyProGlyLeuAlaAla 60
        |||||
DB      2127 GAGCGGCGGAGCGGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 2186

QY      61 AlaLeuArgAspAlaAspLeuValLeuLeuAlaValHisGlyAspValAlaLeuValGAla 80
        |||||
DB      2187 GCCCTCCGGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 2246

QY      81 ValAlaProValThrArgLeuMetArgProGlyValAlaLeuLeuAlaAspThrLeuSerVal 100
        |||||
DB      2247 GTGGCGGCTGAGCCCGGCTCATGCGAGCGGCGGCGGCTGCTCCCGAGACCTGTCGCTC 2306

QY      101 ArgThrGlyMetAlaAlaGlyLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeu 120
        |||||
DB      2307 CGGACGGGAGTGGCGGCGGAGCTCGCGGCCACGCCCGCGGCTCCAGGAGCGGAGCGGAGCTC 2366

QY      121 AsnProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValValThr 140
        |||||
DB      2367 AACCCGATGTTCCCGCCCGCGGCGGAGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2426

QY      141 ArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGlyGlyGlyArgProPro 160
        |||||
DB      2427 AGGAGCGGCGGCGGCGGCTGCGGCGGCTGCGGCTGCTGAGGAGCGGCGGCGGAGGCGCC 2486

QY      161 ValArgLeuThrAlaGlyGlyHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHis 180
        |||||
DB      2487 GTACGGCTGACGGCGGAGGAGGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 2546

QY      181 AlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValAlaGAlaLeu 200
        |||||
DB      2547 GCGGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2606

QY      201 AlaAlaThrAlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGly 220
        |||||
DB      2607 GCGGCGGAGCGGAGCGGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 2666

QY      221 GlySerProGlyValArgGlyAspAlaGlnArgSerAsnProArgAlaAlaSerAlaArg 240
        |||||
DB      2667 GCGAGCGGCGGAGGAGTACCGGAGCATCCAGCGGTCCAGCCCGGCGGCGGCTCGCGCGCG 2726

QY      241 ArgAlaLeuAlaGlyAlaLeuArgSerPheAlaAlaLeuValGlyAsp----- 256
        |||||
DB      2727 CCGGCGCTCGCGCGAGGCGCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2786

QY      257 -----AspProAspArgAlaAspAlaProGlyArgAlaAspAlaProGlyHisProGly 274
        |||||
DB      2787 GCCGAGGAGCGGAGCGGCGGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 2846
```

QY 275 GlyCysAspGlyAlaGlyAsnLeuAspGlyValPheGlyGlyLeuArgArgLeuMetGly 294

DB 2847 GGATGCGAGCGGCGGCGGAGAACCTCCGAGCGGCTGCTGAGGAACTCCGCGGCTCATGGGA 2906

QY 295 ProGlyLeuAlaAlaGlyGlnAspHisCysGlnGlyLeuPheArgThrLeuHisArgThr 314

DB 2907 CCGGAGCTCGCGGCGGCGGAGGAGCACTCCAGGAGCTGTCGACCTCCACCGCACC 2966

QY 315 AspAspGlyGlyGlyLeuAspArg 322

DB 2967 GACGAGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2990

## RESULT 4

ABZ69798  
ID ABZ69798 standard; DNA; 12391 BP.

AC ABZ69798;

DT 08-APR-2003 (first entry)

DE Plasmid plasc-papabc.

KW DHFR, translation; orthogonal tRNA, O-tRNA; O-RS; CAT; TyrRS;

KW orthogonal aminoacyl tRNA synthetase; unnatural amino acid;

KW chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;

XX biosynthesis; p-aminophenylalanine; pAF.

XX Synthetic.

XX WO200285923-A2.

PF 19-APR-2002; 2002WO-US012465.

PR 19-APR-2001; 2001US-0285030P.

PR 06-FEB-2002; 2002US-0355514P.

XX (SCRI ) SCRIPFS RES INST.

PI Schultz P, Wang L, Anderson JC, Chin JW, Liu DR, Magliery TJ;

PI Meggers EL, Nehl RA, Pasternak M, Santoro SW, Zhang Z;

XX WPI; 2003-120430/11.

PT Composition useful for producing protein comprising unnatural amino acid,

PT has translation system comprising orthogonal tRNA and orthogonal

XX aminoacyl tRNA synthetase.

PS Example 4; Page 124-127; 1889p; English.

XX The invention relates to a novel composition comprising a translation

CC system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl

CC tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O

CC tRNA with at least one unnatural amino acid in the translation system

CC and the O-tRNA recognises at least one selector codon. A composition of

CC the invention is useful for producing at least one protein comprising at

CC least one unnatural amino acid. The protein is the Asp127G mutant of

CC chloramphenicol acetyltransferase (CAT), the Tyr163Tg mutant of mouse

CC dihydrofolate reductase (DHFR), or a Tyr163Tg mutant of mouse

CC acid is provided exogenously. The translation system is a cell and the

CC unnatural amino acid is biosynthesised by the cell. The present sequence

CC represents a plasmid for use in the biosynthesis of p-aminophenylalanine

XX (pAF) in vivo

XX Sequence 12391 BP; 2830 A; 3588 C; 3346 G; 2627 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	5,16e-84	Length:	12391
Score:	1588.00	Matches:	315
Percent Similarity:	96.65%	Conservative:	2
Best Local Similarity:	96.04%	Mismatches:	5



Query Match: 97.13% Indels: 6  
 DB: 8 Gaps: 1  
 US-10-089-514-6 (1-322) x ABZ69798 (1-12391)

QY 1 MetSerGlyPheProArgSerValValaIGlySerGlyAlaValaIGlyMetPhe 20  
 DB 2361 ATGATGGCTTCCCGGAGCGTGTGTGGGGGAGGAGCGGTGGGCGATGTTTC 2420

QY 21 AlaGlyLeuLeuArgGlyAlaGlySerArgThrLeuValValaPheLeuValProProPro 40  
 DB 2421 GCGGGGTGTGTGGGGAGGCGGAGCGGACCGCTCTGTCCAGCTCGTACCGCGCGCG 2480

QY 41 GlyArgProAspAlaCysLeuValaGlyAspValThrAlaProGlyProGlyLeuAlaAla 60  
 DB 2481 GAGAGCGCGGACCGCTGTGTGTGGGAGCGTACCGCGCGGCGCGGCTGTGGGGCC 2540

QY 61 AlaLeuArgAspAlaAspLeuValaLeuLeuAlaValaIHisGlyAspValaAlaLeuVala 80  
 DB 2541 GCGCTCGGGAGCGGAGCTGT 2600

QY 81 ValaIleProValaThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerVal 100  
 DB 2601 GTGGCGCGCGT 2660

QY 101 ArgThrGlyMetAlaIleGlyLeuAlaAlaIleAlaProGlyValaGlyIleValaGlyLeu 120  
 DB 2661 CGGACGGGAGTGGCGGGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2720

QY 121 AspProMetPheAlaProAlaAlaGlyMetThrGlyArgProValaAlaAlaValaThr 140  
 DB 2721 AACCCATGTTCGCCCGCGCGCGGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2780

QY 141 ArgAspGlyProGlyValaThrAlaLeuLeuArgLeuValaGlyGlyGlyArgPro 160  
 DB 2781 AGGAGCGGCGCGCGCGT 2840

QY 161 ValArgLeuThrAlaGlyIleHisAspArgThrThrAlaAlaThrGlyAlaLeuThrHis 180  
 DB 2841 GTACGGCTCACGGGAGGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 2900

QY 181 AlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValaAspValaArgAlaLeu 200  
 DB 2901 GCGGTATCT 2960

QY 201 AlaAlaThrAlaProProProHisGlyValaLeuLeuAlaLeuLeuAlaArgValaLeuGly 220  
 DB 2961 GCGGCGAGCGGACCGCGCGCGGACCGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3020

QY 221 GlySerProGlyValaThrGlyAspIleGlyArgSerAspProArgAlaAlaSerAlaArg 240  
 DB 3021 GGGAGCGCGGAGGT 3080

QY 241 ArgAlaLeuAlaGlyAlaLeuArgSerPheAlaAlaLeuValaGlyAsp----- 256  
 DB 3081 CGGGCGCTCGCGCGCGCGCT 3140

QY 257 -----AspProAspArgAlaAspAlaProGlyArgAlaAspAlaProGlyHisProGly 274  
 DB 3141 GCGGAGGAGCGGAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 3200

QY 275 GlyCysAspGlyValaGlyLeuLeuAspGlyValaPheGlyGlyLeuArgGlyLeuMetGly 294  
 DB 3201 GAGTGTGAGCGGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 3260

QY 295 ProGlyLeuAlaAlaGlyGlyAlaAspHisGlyGlyGlyLeuLeuPheArgThrLeuHisArgThr 314  
 DB 3261 CGGAGCTCGCGCGCGCGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 3320

QY 315 AspAspGlyGlyGlyValaAspArg 322  
 DB 3321 GACGAGGAGCGGAGGAGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3344

RESULT 5  
 AAT59269  
 ID AAT59269 standard; cDNA; 888 BP.  
 AC AAT59269;  
 AC AAT59269;  
 DT 02-APR-1997 (first entry)  
 XX  
 DE Streptomyces pristinaespiralis papC gene.  
 XX  
 KM Streptogramin B; antibiotic; biosynthesis; pristinamycin; virginiamycin;  
 KM DMPAPa precursor; 4-dimethylamino-L-phenylalanine; papA; papM; papB;  
 KM papC; isomerisation; aromatisation; N-methyltransferase; ds.  
 OS Streptomyces pristinaespiralis.  
 XX  
 FH Key Location/Qualifiers  
 FT mat\_peptide 1..888  
 FT /\*tag= a  
 FT /product= "PapC"  
 XX  
 PN MO9601901-A1.  
 XX  
 PD 25-JAN-1996.  
 XX  
 PF 04-JUL-1995; 95MO-FR000889.  
 XX  
 PR 08-JUL-1994; 94FR-00008478.  
 XX  
 PA (RHON ) RHONE POULENC ROBER SA.  
 XX  
 PI Blanc V, Thibaut D, Bamas-Jacques N, Blanche F, Cruzet J;  
 PI Barriere J, Debussche L, Famechon A, Paris J, Dutruc-Roseet G;  
 DR MPI; 1996-097631/10.  
 XX  
 PT New streptogramin B derivs. useful as antibiotics - produced by new  
 PT mutants of Streptomyces having altered genes for streptogramin B  
 PT biosynthesis.  
 XX  
 PS Claim 18; Page 105-106; 146pp; French.  
 XX  
 CC The papA and papM genes of S. pristinaespiralis are involved in the  
 CC biosynthesis of 4-dimethylamino-L-phenylalanine (DMPAPa), a precursor for  
 CC pristinamycin IA. The region between these two genes was sequenced and  
 CC two open reading frames were identified. The first (papC) was on the  
 CC complementary strand and decodes to an amino acid sequence with homology  
 CC to the region of E. coli TyrA which has been implicated in aromatisation  
 CC reactions. The papC gene product is likely to be involved in a similar  
 CC phenylpyruvate during DMPAPa synthesis. The second open reading frame  
 CC (papB) could be decoded to give a product with homology to the region of  
 CC TyrA which has chorismate mutase activity. The papB gene product is  
 CC likely to catalyse the isomerisation of 4-deoxy 4-amino chorismate to 4-  
 CC deoxy 4-aminoprephenate acid. Disruption of the papB and papC genes can  
 CC produce the antibiotic strains of S. pristinaespiralis which are unable to  
 CC new, modified forms of it  
 XX  
 SO Sequence 888 BP; 73 A; 254 C; 428 G; 133 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 1.41e-30 Length: 888  
 Score: 661.50 Matches: 155  
 Percent Similarity: 60.38% Conservative: 34  
 Best Local Similarity: 49.52% Mismatches: 101  
 Query Match: 40.46% Indels: 23  
 DB: 2 Gaps: 4  
 US-10-089-514-6 (1-322) x AAT59269 (1-888)  
 QY 2 SerGlyPheProArgSerValValaIGlySerGlyAlaValaIGlyMetPheAla 21



```

DB      13 TCAGTGTCTGGGCGTGTGTGTGTGTGGCGGGCGGTGCGGCGCATGTTCAAC 72
QY      22 GtyleuLeuAargtluAglYserArgThrleuValValAspIleuValProProProglY 41
DB      73 CACTGCGTGTGGGTGGGGGTGGCGGTGACCTGCGTGCAC--GTGGCGGGGCGCGGT 129
QY      42 ArgProAspAlaCysIleuVal-----GtYAspValThrAlaProglYProglIleuAla 59
DB      130 GCGCGGACGCGGGGTGGCGGTGGCGGTGATGTGCGGCGCGCGGGCGCGAGGGCGTCC 189
QY      60 AlaAlaIleuAArgAspAlaAspIleuValIleuLeuAlaValHisGluAspValAlaIleuLys 79
DB      190 GCGCGCTGGCGCGCGCGGACGTGTGTGTGTGCGCGGTGCGCGGCGCGGTGGCGGTGGAG 249
QY      80 AlaValAlaProValThrArgIleuMetArgProglYAlaIleuLeuAlaAspThrIleuSer 99
DB      250 GCGGTGTGAGTGTGTGGCGGGGTGATGTGGCGCGGTGCGGTGCGGTGCGGTGCGGTGCG 309
QY      100 ValArgThrGtYMetAlaIaIaIaIeulAlaHisAlaProglYValGlnHisValGtY 119
DB      310 GTCAAGAGCGCGGATTCGCGCGCGCGGTGCGAGCGCGCGCGGCGTGCAGCGGTGGGG 369
QY      120 LeuAsnProMetPheAlaProAlaAlaGtYMetThrGtYArgProValAlaAlaValVal 139
DB      370 CTGAACCCGATGTTCCGCCCTGCGGTGGTCTTCAAGGGGCGCGCGGTGGCGGTGGTGTG 429
QY      140 ThrArgAspGtYProglYValThrAlaIleuLeuAArgIleuValGluGtYGtYArg 159
DB      430 GTCAACGACGCGCGCGGTGTGGCGCGGTGTGTGACGTGTGGCGCGGTGGCGGGCGCGG 489
QY      160 ProValArgIleuThrAlaGluGlnHisAspArgThrThrAlaIaThrGlnAlaIeulThr 179
DB      490 GTGTGTGAAGATGCGCGCGCGCGGCAAGACGACGTACCGCGCGAGCGAGCGCGCGCAG 549
QY      180 HisAlaValIleuLeuSerPheGtYLeuAlaIleuAlaArgIleuGtYValAspValArgAla 199
DB      550 CATGCGCGCGGTGCTGTGCGGTGTGGCGGTGTGGCGGTGTGTGCGGTGTGCGGTGTGCG 609
QY      200 IeuAlaAlaThrAlaProProProHisGlnValIleuLeuAlaIleuLeuAlaArgValIeu 219
DB      610 CTGCGGACAGTCCCGCGCGCGCATGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 669
QY      220 GtYGtYSerProGluValTyrGtYAspIleuGlnArgSerAsnProArgAlaAlaSerAla 239
DB      670 GCGGGACGCGCGAGGTGATTTGCAATTCACAGCGCGCAACCGCGCGCGCGCGCGCGG 729
QY      240 ArgArgAlaIleuAlaIaIeulAlaIeulArgSerPheAlaAlaIeulValGtYAspAspProAsp 259
DB      730 CCGACAGCGCGTGGCGCGCGCGCTG-----GTGGCGGTGGGGCGAGCGCGTGTGAG 777
QY      260 ArgAlaAspAlaProglYArgAlaAspAlaProglYHisProglYGtYAspArgValAla 279
DB      778 AGGGGCGACGAGAGACGTTGCGC----- 801
QY      280 GtYAsnIleuAspGtYValPheGtYGluIleuAArgIleuMetGtYProglIleuAlaAla 299
DB      802 -----GCCCTGTTCCGCCGAATCCGCGGTGTGTGGCGAGACGCGCGCGAG 849
QY      300 GtYGlnAspHisCysGlnGluIleuPheArgThrIleuHis 312
DB      850 CTGGAACGCGTGTGGCGCGAGTGTTCACCGCCCTGTGCAC 888

```

RESULT 6  
AAT59268/c  
ID AAT59268 standard; cDNA; 2888 BP.

XX AC AAT59268;  
XX DT 02-APR-1997 (first entry)  
XX DE Streptomyces pristinaespiralis papA and papM intergenic region.  
XX

```

KM Streptogramin B; antibiotic; biosynthesis; pristinamycin; virginiamycin;
KM DMPAPA precursor; 4-dimethylamino-L-phenylalanine; PAPa; papM; papB;
KM papC; isomerisation; aromatisation; N-methyltransferase; ds.
OS Streptomyces pristinaespiralis.
XX
XX Key Location/Qualifiers
XX CDS 1..687
XX FT /tag= a
XX FT /product= "PAPA"
XX FT /note= "C-terminal coding region only, i.e. a partial
XX CDS complement (949..1836)
XX FT /tag= b
XX FT /product= "papC"
XX CDS 1873..2262
XX FT /tag= c
XX FT /product= "PAPB"
XX CDS 2259..2888
XX FT /tag= d
XX FT /product= "PAPM"
XX FT /note= "N-terminal coding region only, i.e. a partial
XX open reading frame"
XX
XX MO9601901-A1.
XX
XX 25-JAN-1996.
XX
XX 04-JUL-1995; 95WO-FR000889.
XX
XX 08-JUL-1994; 94FR-00008478.
XX
XX (RHON ) RHONE POULENC RORER SA.
XX
XX Blanc V, Thibaut D, Bamas-Jacques N, Blanche F, Crouzet J;
XX Barriere J, Debussche L, Famechon A, Paris J, Dutruc-Kosset G;
XX
XX MPI; 1996-097631/10.
XX P-PSDB; AAM11582, AAM11583.
XX
XX New streptogramin B derivs. useful as antibiotics - produced by new
XX mutants of Streptomyces having altered genes for streptogramin B
XX biosynthesis.
XX
XX Example 1; Page 102-104; 146gp; French.
XX
XX PS The PAPa and PAPM genes of S.pristinaespiralis are involved in the
XX CC biosynthesis of 4-dimethylamino-L-phenylalanine (DMPAPA), a precursor for
XX CC pristinamycin IA. The region between these two genes was sequenced and
XX CC two open reading frames were identified. The first (papC) was on the
XX CC complementary strand and decodes to an amino acid sequence with homology
XX CC to the region of E.coli TyrA which has been implicated in aromatisation
XX CC reactions. The PAPC gene product is likely to be involved in a similar
XX CC aromatisation of 4-deoxy 4-amino prephenate to give 4-amino
XX CC phenylpyruvate during DMPAPA synthesis. The second open reading frame
XX CC (papB) could be decoded to give a product with homology to the region of
XX CC TyrA which has chorismate mutase activity. The PAPB gene product is
XX CC likely to catalyse the isomerisation of 4-deoxy 4-amino chorismate to 4-
XX CC deoxy 4-aminoprephenate. Disruption of the papB and papC genes can be
XX CC used to produce strains of S.pristinaespiralis which are unable to
XX CC produce the antibiotic pristinamycin I but which may be able to produce
XX CC new, modified forms of it
XX
XX SQ Sequence 2888 BP; 390 A; 1319 C; 897 G; 282 T; 0 U; 0 Other;

```

#### Alignment Scores:

Pred. No.: 4.27e-30 Length: 2888  
Score: 661.50 Matches: 155  
Percent Similarity: 60.38% Conservative: 34  
Best Local Similarity: 49.52% Mismatches: 101  
Query Match: 40.46% Indels: 23  
DB: 2 Gaps: 4



```

US-10-089-514-6 (1-322) x AAT59268 (1-2888)
QY      2 SerGlyPheProArgSerValValValGlyGlySerGlyAlaValGlyMetPheAla 21
      1824 TCGGTGTTCCGGGCGGTTGTGTGTGGGGCGGCGCGGTGGGGCCCATGTTCCAGC 1765
QY      22 GlyLeuLeuArgGluValGlySerArgThrLeuValAlaLeuValProProProGly 41
      1764 CACTGCTGTGGCTGTGGGGGTGGCGGTGACCTGGGTGGAC---GTGGCCGGGGCCGGT 1708
QY      42 ArgProAspAlaCysLeuVal-----GlyAspValThrAlaProGlyProGluLeuAla 59
      1707 GCGGCGGACGGGGTGGCGGTGTGGCGGTGATGTGGCGGGCGGGCGGGCGGGCGGTG 1648
QY      60 AlaAlaLeuArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuLys 79
      1647 GCGGCGCTGGCGCGCGCGGACGTGTGTGTCTGTGGCGGTGCGGACGCGGTGGGTGGAG 1588
QY      80 AlaValAlaProValThrArgLeuMetArgProGlyValAlaLeuMetPheLeuSer 99
      1587 GCGGTGAGGTGCTGTGGCGGGGTGATGTGGCGCGGTGTGGCTGTGGCGACCTGTGCG 1528
QY      100 ValArgThrGlyMetAlaAlaGluLeuAlaAlaHisAlaProGlyValGlnHisValGly 119
      1527 GTCAAGACCGCGGATCGCGGGCGGCTGTGTGAGGCGCGCGGGCTGCAGAGGGGTGGGG 1468
QY      120 LeuAspProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValVal 139
      1467 CTGAACCCGATGTTCGCGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1408
QY      140 ThrArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGluGlyGlyValArg 159
      1407 GTCAACGACGGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1348
QY      160 ProValArgLeuThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThr 179
      1347 GTGTGTGAGATGCGCGCGCGCGCGCGACGAGTGTGACCGCGCGCGCGCGCGCGCGCG 1288
QY      180 HisAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAla 199
      1287 CATGCCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1228
QY      200 LeuAlaAlaThrAlaProProProProHisGlnValLeuLeuAlaLeuAlaArgValLeu 219
      1227 CTGCGGACGATCGCGCGCGCGCGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1168
QY      220 GlyGlySerProGluValTyrGlyAspPheGlnArgSerAspProArgAlaAlaSerAla 239
      1167 GGGGCGACGCGCGGAGGTGTATTTGCATTCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1108
QY      240 ArgArgAlaLeuAlaGluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAsp 259
      1107 CGGACGCGCGCTGGCGCGCGCGCTG-----GTGCGGTGGGGCGAGCGCGGTGAG 1060
QY      260 ArgAlaAspAlaProGlyArgAlaAspAlaProGlyHisProGlyGlyCysAspGlyAla 279
      1059 AGGAGCGACGACGAGACGTTCCGCC----- 1036
QY      280 GlyAsnLeuAspGlyValPheGlyGluLeuArgArgLeuMetGlyProGluLeuAlaAla 299
      1035 -----GCCCTGTTCCGCGAATGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 988
QY      300 GlyGlnAspHisCysGlnGluLeuPheArgThrLeuHis 312
      987 CTGAACGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG
DB      987 CTGAACGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG
AC      ACF71774;
XX      20-NOV-2003 (first entry)

```

```

XX      Photobhabdus luminescens nucleotide sequence #10241.
DE      Antibiacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW      detection; food; gene expression; plant; animal; microorganism; toxin;
KW      antibiotic; biopesticide; virulence factor; disease model; plague;
KW      whooping cough; gene; ds.
OS      Photobhabdus luminescens.
PN      WO200294867-A2.
XX      28-NOV-2002.
XX      07-FEB-2002; 2002WO-IB003040.
XX      07-FEB-2001; 2001FR-00001659.
XX      (INSP ) INST PASTEUR.
XX      (CNRS ) CNRS CENT NAT RECH SCI.
PI      Duchaud E, Tacourit S, Glaeser P, Frangeul L, Kunst F, Danchin A,
PI      Buchrieser C;
DR      WPI; 2003-148459/14.
XX      Genomic sequence of Photobhabdus luminescens and encoded polypeptides,
PT      useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX      Claim 2; SEQ ID NO 10241; 1205bp; French.
XX      The invention relates to the isolation of genes and their encoded
XX      proteins from Photobhabdus luminescens. The isolated sequences are
XX      sources of probes and primers for detecting the genome of P. luminescens
XX      and related species; to study polymorphisms; for gene analysis and for
XX      detection/amplification of the genes. Antibodies (Ab) raised against the
XX      polypeptides encoded by the genes are used for detection/identification
XX      of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
XX      carry a gene-containing vector are used to select compounds that
XX      modulate, regulate, induce or inhibit expression of the genes in plants,
XX      animals or microorganisms other than P. luminescens and are able to alter
XX      response or sensitivity to toxins and antibiotics produced by P.
XX      luminescens. Cells transformed to express the genes are useful for
XX      recombinant production of the proteins, particularly toxins and
XX      antibacterials useful as insecticides, bactericides and fungicides. The
XX      genes, proteins, vectors containing the genes and Ab are also useful
XX      therapeutically (to treat microbial infection by bacteria or fungi that
XX      are sensitive to P. luminescens-encoded toxins or antibiotics) and as
XX      biopesticides. Other uses of the genes and the proteins are as virulence
XX      factors and for identifying targets of human diseases for which P.
XX      luminescens is a model (particularly plague and whooping cough). This
XX      sequence represents one of the isolated P. luminescens genes
XX      SQ      Sequence 888 BP; 207 A; 194 C; 229 G; 258 T; 0 U; 0 Other:
XX      Alignment Scores:
XX      Pred. No.: 3,92e-14 Length: 888
XX      Score: 382.00 Matches: 101
XX      Percent Similarity: 47.94% Conservative: 50
XX      Best Local Similarity: 32.06% Mismatches: 126
XX      Query Match: 23.36% Indels: 38
XX      DB: 10 Gaps: 5

```

```

US-10-089-514-6 (1-322) x ACF71774 (1-888)
QY      8 ValValValGlyGlySerGlyAlaValGlyMetPheAlaGlyLeuLeuArgGluAla 27
      19 GTATCTCTGTGTGGGCAAGCGCTATTGTTCTTTATGAGCGCTTTTACTCAATAT 78
DB      28 GlySerArgThrLeuValAspLeuValProProProGlyArg----- 42
      79 GGTCAATTGTATACAGTGTGTGCAACGACGAGAGTGTGTCTCATTTACTATCAT 138

```



WP	Fragment Name	Begin	End	LOCUS	ACF67367	Accession	Actf67367
WP	ACF67367_01	1	110000				
WP	ACF67367_02	200001	310000				
WP	ACF67367_03	300001	410000				
WP	ACF67367_04	400001	510000				
WP	ACF67367_05	500001	610000				
WP	ACF67367_06	600001	710000				
WP	ACF67367_07	700001	810000				
WP	ACF67367_08	800001	910000				
WP	ACF67367_09	900001	1010000				
WP	ACF67367_10	1000001	1110000				
WP	ACF67367_11	1100001	1210000				

WP	ACF67367_12	120001	111000					
WP	ACF67367_13	130001	141000					
WP	ACF67367_14	140001	151000					
WP	ACF67367_15	150001	161000					
WP	ACF67367_16	160001	171000					
WP	ACF67367_17	170001	181000					
WP	ACF67367_18	180001	191000					
WP	ACF67367_19	190001	201000					
WP	ACF67367_20	200001	211000					
WP	ACF67367_21	210001	221000					
WP	ACF67367_22	220001	231000					
WP	ACF67367_23	230001	241000					
WP	ACF67367_24	240001	251000					
WP	ACF67367_25	250001	261000					
WP	ACF67367_26	260001	271000					
WP	ACF67367_27	270001	281000					
WP	ACF67367_28	280001	291000					
WP	ACF67367_29	290001	301000					
WP	ACF67367_30	300001	311000					
WP	ACF67367_31	310001	321000					
WP	ACF67367_32	320001	331000					
WP	ACF67367_33	330001	341000					
WP	ACF67367_34	340001	351000					
WP	ACF67367_35	350001	361000					
WP	ACF67367_36	360001	371000					
WP	ACF67367_37	370001	381000					
WP	ACF67367_38	380001	391000					
WP	ACF67367_39	390001	401000					
WP	ACF67367_40	400001	411000					
WP	ACF67367_41	410001	421000					
WP	ACF67367_42	420001	431000					
WP	ACF67367_43	430001	441000					
WP	ACF67367_44	440001	451000					
WP	ACF67367_45	450001	461000					
WP	ACF67367_46	460001	471000					
WP	ACF67367_47	470001	481000					
WP	ACF67367_48	480001	491000					
WP	ACF67367_49	490001	501000					
WP	ACF67367_50	500001	511000					
WP	ACF67367_51	510001	521000					
WP	ACF67367_52	520001	531000					
WP	ACF67367_53	530001	541000					
WP	ACF67367_54	540001	551000					
WP	ACF67367_55	550001	561000					
WP	ACF67367_56	560001	564894					

Alignment Scores:  
Pred. No.: 3.67e-12 Length: 110000  
Score: 382.00 Matches: 101  
Percent Similarity: 47.94% Conservative: 50  
Best Local Similarity: 32.06% Mismatches: 126  
Query Match: 23.36% Indels: 38  
DBs: 10 Gaps: 5

US-10-089-514-6 (1-322) x ACF67367\_49 (1-110000)

QY	8 ValValValGlyGlyGlySerGlyAlaValAlaGlyGlyMetPheAlaGlyPheLeuAlaArgGluAla 27
Db	76463 GTCATTCCTGGTGGGCAAGCGCGGATTTGGTTCTTATTTAGAGCGCCTTTTACTCAATAT 76522

QY	28 GlySerArgThrLeuValAlaAspLeuValProProProGlyArg----- 42
Db	76523 GGTCAATTTGTAATCACTGTTCACAAACGACGAGAGTGGCTTCTTCATTATCAT 76582

QY	43 ProAspAlaCysLeuValGlyAspValThrAlaProGlyProGlyLeuAlaAlaAlaLeu 62
Db	76583 TCTCATTAAGTGTGAGATT---GATATCTTAAACCGCGTGCATGTGTGGACGGTTTGG 76639

QY	63 ArgAspAlaAspLeuValLeuLeuAlaValAlaHisGlyAspValAlaLeuValAlaAla 82
Db	76640 AGTGGGCAAGGCGGTCCTTTCCTTAACCGAAGCCGTGACCATACAGCGTTGGCT 76699

83 ProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThr 102







WP ADM27081\_07 700001 810000  
 WP ADM27081\_08 800001 910000  
 WP ADM27081\_09 900001 1010000  
 WP ADM27081\_10 1000001 1110000  
 WP ADM27081\_11 1100001 1210000  
 WP ADM27081\_12 1200001 1310000  
 WP ADM27081\_13 1300001 1410000  
 WP ADM27081\_14 1400001 1510000  
 WP ADM27081\_15 1500001 1610000  
 WP ADM27081\_16 1600001 1694968

## Alignment Scores:

Pred. No.: 1.09e-08 Length: 110000  
 Score: 323.00 Matches: 93  
 Percent Similarity: 52.55% Conservative: 41  
 Best Local Similarity: 36.47% Mismatches: 104  
 Query Match: 19.76% Indels: 18  
 DB: 11 Gaps: 8

US-10-089-514-6 (1-322) x ADM27081\_07 (1-110000)

QY 6 ArgSerValValAlaGlyGlySerGlyAlaValAlaGlyMetPheAlaGlyLeuArg 25  
 DB 59054 AGGATCGCGATCTCGCGGCGGCGCATGAGCGCTTATCGCCCGGAGCTCCGT 59113  
 QY 26 GluAlaGly-SerArg-----ThrLeuValAlaAspLeuValProProGlyArgPr 43  
 DB 59114 GAGGACGGTCATAGAGTACCTACTGTTCAACCTCAGACCCGCGGAGCGCTCGCC 59173  
 QY 43 oAspAlaCysLeuValGlyAspValThrAlaProGlyProGlyLeuAlaAlaLeuAr 63  
 DB 59174 CGCGAGCTC-----GACCTCGAGGCT--GCTCCGACCAAGCTGATCGCCCAA 59220  
 QY 63 gAspAlaAspLeuValLeuLeuAlaAlaHisGluAspValAlaLeuValAlaAlaPr 83  
 DB 59221 AGATGCCGACGTTGTGTTGTTCCGTCCTCCCATTCGCTACGAAAGCTCATCCGTA 59280  
 QY 83 oValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrG 103  
 DB 59281 GGTTCGACCAACAGCTTCCTGAGGTTCCCTCTCACCAGCACTCCCTCAAGGTTCG 59340  
 QY 103 yMetAlaAlaGluLeuAlaAlaHisAlaPro--GlyValGlnHisValGlyLeuAsnPr 122  
 DB 59341 TCCCGTTCGCGTATGCTCGAGCAGCTCCGAGAGATGTGTACGTCTCGAAGCTCACC 59400  
 QY 122 oMetPheAlaProAlaAla--GlyMetThrGlyArgProValAlaAlaValAlaThr-- 140  
 DB 59401 CTTCTTCGCGCCACGCGTCCCGAGTCTCCGCGGTCAAGCCGTATCTCTCAAGCCGGA 59460  
 QY 141 -ArgAspGlyProGlyValThrAlaLeuArgLeuValGluGlyGlyValArgPr 160  
 DB 59461 GCGTACTGTCCTCGAGCTCGCGCGCTGCTCTGATCTGAGACGGAAGCGCTGAGT 59520  
 QY 160 oValArgLeuThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrH 180  
 DB 59521 CGTTGAGACCAACCCCGAGAAACACATCGCACTATGCGCGTTCAGTGCTCACCCA 59580  
 QY 180 sAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArg-----LeuGlyValAs 196  
 DB 59581 TGGTGTCTCTCTGTCGACGCGCCGCAATGAGCCGCTTCTCCCAAGTTCAGAGCTCA 59640  
 QY 196 pValArgAlaLeuAlaAlaThrAlaProProHisGlnValLeuLeuAlaLeuLeuAl 216  
 DB 59641 CATCGAAGAGTGCACG-----CCCGTATCGCTTACTATGAGACGTCTGCGG 59691  
 QY 216 aArgValLeuGlyGlySerProGluValArgGlyAspIleGlnArgSerAsnProArgAl 236  
 DB 59692 TCGTATCGCGGACAGATCAGTCTTTACGCGGAGATCAAGCTTCAACCCCTTACGG 59751  
 QY 236 aAlaSerAlaArgArgAlaLeuAlaGluAlaLeuArgSerPhe 250  
 DB 59752 TGACGAAGCCCGTGAAGAGCTCTCCGCGCCCTCAAGCGCTTC 59794

## RESULT 11

AAD50753

ID AAD50753 standard; DNA; 1122 BP.

AC AAD50753;

DT 02-APR-2003 (first entry)

DE Escherichia coli bifunctional prephenate dehydrogenase DNA.

KW Tocopherol biosynthesis; prephenate dehydrogenase; chorismate mutase;  
 KW tyra; transgenic; transgenic plant; homogenetic acid; plastoquinone;  
 KW enzyme; gene; ds.

XX Escherichia coli.

FH Key Location/Qualifiers

FT CDS 1..1122  
 FT /tag= a  
 FT /product= "Bifunctional prephenate dehydrogenase"

XX MO200289561-A1.

XX 14-NOV-2002.

XX 03-MAY-2002; 2002WC-US013898.

XX 09-MAY-2001; 2001US-0289527P.

XX (MONS ) MONSANTO TECHNOLOGY LLC.

XX Valentin HE, Miley TA;

XX MPI; 2003-058889/05.

XX P-PSDB; AAB32998.

XX Novel nucleic acid molecule encoding enzymes involved in tocopherol  
 PT biosynthesis useful for enhancing or altering tocopherol production in  
 PT plants such as canola, maize, soybean, crame, mustard, castor bean,  
 PT peanut, sesame, or cottonseed.

XX Claim 5; Page 201-202; 206pp; English.

XX The present invention relates to novel nucleic acid molecules encoding  
 CC enzymes involved in tocopherol biosynthesis. Polynucleotides of the  
 CC invention comprise a promoter region functioning in a plant cell to cause  
 CC production of an mRNA molecule linked to a heterologous nucleic acid  
 CC molecule encoding an enzyme with prephenate dehydrogenase (tyra) or  
 CC chorismate mutase activities or a tyra from Erwinia herbicola or  
 CC Escherichia coli bifunctional prephenate dehydrogenase. Sequences of the  
 CC invention are useful for producing a plant such as canola, maize,  
 CC Arabidopsis, Brassica campestris, B.mapus, soybean, crame, mustard,  
 CC castor bean, peanut, sesame, cottonseed, linseed, safflower, oil palm,  
 CC flax and sunflower having increased tocopherol levels, especially alpha-  
 CC or gamma-tocopherol or tocotrienol relative to a plant with a similar  
 CC genetic background but lacking the exogenous nucleic acid molecule. They  
 CC are useful for reducing tocopherol levels in a plant. They are also  
 CC useful for producing plant expressing homogenetic acid and  
 CC plastoquinones. The present sequence is Escherichia coli bifunctional  
 CC prephenate dehydrogenase DNA

XX SO Sequence 1122 BP; 248 A; 242 C; 339 G; 293 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 1.66e-05 Length: 1122  
 Score: 237.00 Matches: 78  
 Percent Similarity: 45.39% Conservative: 50  
 Best Local Similarity: 27.66% Mismatches: 110  
 Query Match: 14.50% Indels: 44  
 DB: 8 Gaps: 8

US-10-089-514-6 (1-322) x AAD50753 (1-1122)



QY 8 ValValValGlyGlySerGlyAlaValGlyGlyMetPheAlaGlyLeuLeuMetArgGlyAla 27  
 Db 304 GTTATCTGTGGCGCGCGCGCATGATGGAGCGCTGTTCGAGAAAGTGTGACACTCTCG 363  
 QY 28 GlySerArgThrLeuValValAspLeuValProProGlyArgProAspAlaCysLeu 47  
 Db 364 GGTATTACAGCTCCGCGATTCTGGAG-----CAACATGACTGGGATCCGA 405  
 QY 48 ValGlyAspValThrAlaProGlyProGlyLeuAlaAlaLeuArgAspAlaAspLeu 67  
 Db 406 GCGCGGTGATATTGTGGCC-----GATGCCGGAATG 435  
 QY 68 ValLeuLeuAlaValAlaHisGlyAspValAlaAlaLeuValAlaAlaProValThrArgLeu 87  
 Db 436 GTGATGTGTAAGTGTGCCAATCCACGTTACTGACGAAGTAATTTGGCAAAATTACCGCCTTA 495  
 QY 88 MetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrGly---MetAlaAla 106  
 Db 496 CCGAAA--GATTGTAATCTGTGTTGATCTGGCATCTGTAATAAATGACCATTTACAGGCC 552  
 QY 107 GlyLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeuAsnProMetPheAlaPro 126  
 Db 553 ATGCTGTGGCGCGCCACGATGGCCCGTA-----CTGGGTTACACCCAAATGTTCCGTCG 606  
 QY 127 AlaAlaGlyMetThrGlyArgProValAlaAlaValAlaThrArgAspGlyProGlyVal 146  
 Db 607 GACACGGGTAGCGCTGGCAAGCAAGTT-----GTGCTGTGGTGTATGAGCGTAACCG 660  
 QY 147 ThrAla-----LeuLeuArgLeuValGlnGlyGlyGlyArgProValArgLeu 163  
 Db 661 GAGCATTCACCAATGCTTTCTGGAGCAAAATTCAGGTCTGGCGCGCTGTGCATCGTATT 720  
 QY 164 ThrAlaGlnGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeu 183  
 Db 721 AGCGCGGTGACGACGATCAGAAATATGCGGTTATTACGACCTGGCCACTTGTGTACT 780  
 QY 184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaAlaThr 203  
 Db 781 TTGCTGTACGGGCTGCACCTGCAGCAAGAAATGTTCACTTGAGCAACTTCTGGCGCTC 840  
 QY 204 AlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGlyGlySerPro 223  
 Db 841 TCTTCGCGCATTTTACCGCTTGAGCTGGGATGTGTGGCGCACTGTTCTCAGGATCCG 900  
 QY 224 GlyValTyrGlyAspIleGlnArgSerAsnProArgAlaAlaSerAla----- 239  
 Db 901 CAGCTTATGCCGACATTATTATGTCGTACAGCGTAATCTGGCGTTAATCAAACTTAC 960  
 QY 240 -----ArgArgAlaLeuAla 244  
 Db 961 TATAAGCTTTGGCGGAGCGCATGTAAGTTCGTGGACAGGCGATTAAGCAGCGTTATT 1020  
 QY 245 GlyAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArgAlaAspAlaPro 264  
 Db 1021 GACAGTTTCCGCAAGGTGAGCAGCTGTTCGGGATTACGCAACAGGTTTTCAGAGTGA 1080  
 QY 265 GlyArg 266  
 Db 1081 AGCCGC 1086

RESULT 12  
 AAV21209\_09/c  
 Continuation (10 of 17) of AAV21209 from base 900001 (Methanococcus jannaschii circular  
 WP Sequence split into 17 fragments LOCUS AAV21209 Accession Aav21209  
 WP Fragment Name Begin End  
 WP AAV21209\_00 1 110000  
 WP AAV21209\_01 100001 210000  
 WP AAV21209\_02 200001 310000  
 WP AAV21209\_03 300001 410000  
 WP AAV21209\_04 400001 510000  
 WP AAV21209\_05 500001 610000  
 WP AAV21209\_06 600001 710000  
 WP AAV21209\_07 700001 810000

WP	AAV21209_08	800001	910000	
WP	AAV21209_09	900001	1010000	!
WP	AAV21209_10	1000001	1110000	
WP	AAV21209_11	1100001	1210000	
WP	AAV21209_12	1200001	1310000	
WP	AAV21209_13	1300001	1410000	
WP	AAV21209_14	1400001	1510000	
WP	AAV21209_15	1500001	1610000	
WP	AAV21209_16	1600001	1664976	

  

Alignment Scores:				
Pred. No.:	0.00164	Length:	110000	
Score:	235.00	Matches:	71	
Percent Similarity:	42.76%	Conservative:	59	
Best Local Similarity:	23.36%	Mismatches:	158	
Query Match:	14.37%	Indels:	16	
DB:	2	Gaps:	5	

  

US-10-089-514-6 (1-322)	x	AAV21209_09 (1-110000)
QY	9	ValValGIgylgysereGlylaValaGIgylgylmetPhealaglYleuLeuArglualagly 28
DB	30033	ATTATTGGAGAGACTAGTGGTTTAGGGAATGCTTTGGCAAGATTTTAAAAATTAAGGG 29974
QY	29	SerArgThrIeuValaValaAspleuValProProGlyArgProAspAlaCysIeuVal 48
DB	29973	TTTAACTGTATAGTACTGGGAGAGATATTGAGAAAGAAAAATGTTAGAAAGAGTTA 29914
QY	49	GIYAspValIThrAlaProGlyProGluLeuAlaAlaLeuArgAspAlaAspleuVal 68
DB	29913	GGG--GTTGAATTTACTTAACAACAACATTGAAGAGTCT--AAAAAGAGATATTGTT 29860
QY	69	LeuLeuAlaValHisGluAspValAlaLeuValaValaIaProValIThrArgIeuMet 88
DB	29859	ATTGTACAGTGTCCAAATAATGTTTACGAAGAGGTTTATAAAGAGTACGCTCTCATGTT 29800
QY	89	ArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrGlyMetAlaIaGluLeu 108
DB	29799	AGGGAAGAGATGTTATTAAATGACATPAACCTCATTTAAAGAGATTCCTCAAAAGCTAG 29740
QY	109	AlaAlaHisAla--ProGlyValaGlnHisValGIYLeuAsnProMetPheAlaProAla 127
DB	29739	GAAGAGCATGTTTAAAGAGGAGGTACAGTATTTCCAACCCCAATGTTCCGCCCATCA 29680
QY	128	AlaGlyMetThrGlyArgProValAlaAlaValaValThr-----ArgAspGly 143
DB	29679	ACACCTTCTTGTATTAAACAAGGTATTATCCCTTCGAAAGCATTAATAACT 29620
QY	144	ProGlyValIThrAlaLeuLeuArgLeuValGluGIYGIYValArgProValArgLeu 163
DB	29619	GAGTGTGTTAAACAAGGTTTATATTTTATAAAGAAAGAGGGCTTAAGGTTATAGTCATC 29560
QY	164	ThrAlaGluGlnHisAspArgThrIThrAlaAlaThrGlnAlaLeuThrHisAlaValLeu 183
DB	29559	CCCTCAAAAAACATGATGAATTAATGAGGGAATGTTCAGAGGTTTGCTACCTCCGCTTT 29500
QY	184	LeuSerHeGlyLeuAlaLeuAlaArgLeuGIYValAspValArgAlaLeuAlaAlaThr 203
DB	29499	ATATCTTTAGAGCAACGTTTAAAAAGAACTCAAGCTGTATATTAAGAGTCACAAAGATT 29440
QY	204	AlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGIYGlySerPro 223
DB	29439	GCCTCCCAATATATACAGTGTGATGATTTCTATCATTTAGGAGAGATTATAGACAGAACTCC 29380
QY	224	GluValIYrGlyAspIleGlnArgSerAsnProArgAlaAlaSerAlaArgAlaLeu 243
DB	29379	TATTTATATGCTGACATCCAAATGTTTAAATCCAAAGATTAAGAGATTATGATAAACCCTTT 29320
QY	244	AlaGluAlaLeuArgSerPheAlaAlaLeuValGIYAspAspProAspArgAlaAspAla 263
DB	29319	ATTAATCATGCTATGAGAAATCAGTGAATTTGTTAAAAATTAAGATTAAGAAAGGTTTGT 29260
QY	264	ProGlyArgAlaAspAlaProGlyHisProGlyGIYCyAspGlyValaGly----- 280



```

Db      29292  AAGATATGAAGAAGCTGCMAAGCATTTTGGTATGAGCGCAAGAGGGGCTTATTTAT 29200
Oy      281  -----AsnLeuaspGlyValaPheGlyGluLeuArgArgLeuMetGly 294
Db      29199  TCAGATTAAGCCGCTGTTTGCATTAAACATCTGAATATGMAAGCTAAATAACTCATTTGCT 29140
Oy      295  ProGluLeuAla 298
Db      29139  AAAGATGTGGCT 29128

RESULT 13
ABN83448
ID      ABN83448  standard; DNA, 1122 BP.
XX
XX      ABN83448;
AC
XX      27-AUG-2002  (first entry)
XX
XX      Escherichia coli prephenate dehydrogenase coding sequence.
DE
XX      Arogenate dehydrogenase; enzyme; herbicide; tyrosine biosynthesis;
KW      prephenate dehydrogenase; gene; ds.
XX      Escherichia coli.
OS
XX      Key      Location/Qualifiers
FH      CDS      1..1122
FT      /*tag= a
FT      /product= "Prephenate dehydrogenase"
FT
XX      WO200246441-A2.
XX      13-JUN-2002.
XX      PD
XX      PE      05-DEC-2001; 2001WO-FR003832.
XX      PR      05-DEC-2000; 2000FR-00015723.
XX      PA      (AVET ) AVENTIS CROSCIENCE SA.
XX      PI      Matrinage M, Rippert P;
XX      DR      WP1; 2002-500449/53.
XX      DR      P-PSDB; ABB83256.
XX      PT      New nucleic acid encoding arogenate dehydrogenase, useful for identifying
XX      PT      herbicides, also related protein and herbicide-tolerant plants.
XX      PS      Disclosure; Page 58-60; 66pp; French.
XX
XX      The present invention relates to arogenate dehydrogenase (ADH) enzymes
XX      (see ABN83439-ABN83443 and ABB83248-ABB83251). The ADH enzymes catalyse
XX      the last stage of the metabolic pathway of tyrosine biosynthesis and are
XX      potential targets of herbicides. The present sequence is the coding
XX      CC      sequence for prephenate dehydrogenase, which was used to illustrate the
XX      invention
XX
XX      SO      Sequence 1122 BP; 243 A; 245 C; 342 G; 292 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      2.86e-05      Length:      1122
Score:      233.00      Matches:      77
Percent Similarity:      45.04%      Conservative:      50
Best Local Similarity:      27.30%      Mismatches:      111
Query Match:      14.25%      Indels:      44
DB:      Gaps:      8

US-10-089-514-6 (1-322) x ABN83448 (1-1122)
Oy      8  ValValaValGlyGlySerGlyAlaValGlyGlyMetPheAlaGlyLeuLeuArgGluAla 27
Db      304  GTTATTCCTCGCGCGGTGCGCTCAGATGGACGCGCTGTTGCAAGAAATGCTGACCTCTCG 363

```

OY		28	G S E R A G T h L e U a l V a l A s P r o V a l P r o P r o G I Y A n g P r o A s P A C y S e u	47
Db		364	G G T A T C A G G T C G G A T T C T G A G -----CAACATGACTGGGATCGA	405
OY		48	V a l G A s P r a l T h r A A P r o G I Y P r o G I U e U a l A a l A e U A r G A s P A A s P r e u	67
Db		406	G C G G C G T A T T G T G T G C -----GATTCGGGAATG	435
OY		68	V a l L e U e U a l A H N H S I A s P r a l A e U Y S A l A A P r o V a l T h r A g L e u	87
Db		436	G T G A T T T A G T G T G C C A T C C A G T T A C T G A G A A G T A T T G G C A A T T C C G C T T T A	495
OY		88	M e t A g r P r o G I Y A l e U e U a l A s P r h i L e U e r V a L a T h n g l ---M e t A A l a	106
Db		496	C C G A A ---G A T T G A T T C T G T G C A T T G C A T C A G A A A A T G G C C A T T C A G G C	552
OY		107	G I U e U a l A a H S A l A P r o G I Y A l G I N H S V a l G I U e A n P r o M e t P h e A A P r	126
Db		553	A T G C G T G G G G C A T G A T G T C C G T G -----C T G G G C T A C A C C G A T T T C G T C G	606
OY		127	A A A A G I Y e t h r G I Y A n g P r o V a l A a l A e U a l T h r A n g A s P r G I Y A l	146
Db		607	G A C A G C G T A C C T G C A A A C A A G T ---G T G C T G T G A T G A G C T A A A C G	660
OY		147	T h r A -----L e U e U A g L e U a l G I N H S G I Y G I Y A n g P r o V a l A r G L e u	163
Db		661	G A G A T A C C A A T G G T T C T G A C A A A T T A G G T C T G G G G C C T G G C T G C A T T A T T	720
OY		164	T h r A G I G I U H H S A S P r G T h r T h r A A A T h r G I N A l e U e U H H S A l A l e U	183
Db		721	A G C G C G T G A C A G A T A A T A T G C G T T A T T C A G G A C T G C G C A C T T T G T A C T	780
OY		184	L e U e r P h e I Y L e U a l e U a l A r G L e U Y A l S P r a l A r G A l e U a l A T h r	203
Db		781	T T T G C T A C G G C T G C A C C T G C A G A A A A T G T C A G C T T G A G C A A C T T G G C G C T C	840
OY		204	A A A P r P r o P r o H I S G I N V a l L e U e U a l e U e U a l A r G V a l L e U G I Y S e r P r	223
Db		841	T C T T C C C G A T T A C C C C T T G A G C T G G C G A T G G T G G G G C A C T G T T G C T C A G A T C G	900
OY		224	G I U A l Y r G I Y A s P r I e G I N A r G S e r A n P r o A r G A l A S e r A -----	239
Db		901	C A G C T T A T G C G A C A T C A T T A T G T C G I C A G G G T A T C T G G G T T A T C A A A C G T T A C	960
OY		240	-----A r G A G A l e U a l	244
Db		961	T A T A G C G T T C G G A G G C A T T G A G T G C T G A G C A G G C G A T A G C A G C C T T A T T	1020
OY		245	G I U A l e U A r G S e r P h e A l A l e U V a l G I Y A s P r P r o A P r A G A l A S P r A P r	264
Db		1021	G A C G T T T C C G A A G T G G A C C T G T T C G G C A T T A C G C A C A G C G T T T C A G A G T G A A	1080
OY		265	G I Y A n g	266
Db		1081	A G C G C	1086
<hr/>				
RESULT 14				
ABA99612				
ID	ABA99612 standard; DNA; 1238 BP.			
XX				
XX	ABA99612;			
DT	17-MAY-2002 (first entry)			
DE	E. coli tyra DNA.			
XX				
XX	Shikimate pathway; chorismate mutase; tyra; prephenate dehydrogenase;			
KM	vitamin E; vitamin K; ubiquinone; fat-soluble antioxidant; transgenic;			
KW	feed; pharmaceutical; cosmetic; aromatic amino acid; salicylic acid;			
KM	folic acid; phenylpropanoid; flavanoid; plant; abiotic stress resistance;			
XX	tocopherol; tocotrienol; gene; de.			



```

OS Escherichia coli.
XX
XX Key Location/Qualifiers
FT CDS 25..1146
FT /*tag= a
FT /product= "tyrA"
XX
XX MO200200901-A1.
XX
XX 03-JAN-2002.
XX
XX 28-JUN-2001; 2001MO-EP007391.
XX
XX 29-JUN-2000; 2000DE-01030647.
XX 21-DEC-2000; 2000DE-01064454.
XX
XX (SUNG-) SUNGENE GMBH & CO KGAA.
XX
XX Badur R, Geiger M, Kunze I, Sommer S;
XX WPI; 2002-164442/21.
XX P-PSDB; AAM49653.
XX
XX Preparing fine chemicals, particularly Vitamins E and K, useful as
XX antioxidants e.g. in foods or medicine, by growing organisms with altered
XX shikimate biosynthesis pathway.
XX
XX Claim 20; Page 55-57; 74pp; German.
XX
XX This invention describes a novel method for the preparation of fine
XX chemicals by culturing organisms in which the shikimate pathway has been
XX altered relative to the wild type. The method involves a construct
XX containing a plasmid transit peptide, a chorismate mutase and/or
XX prephenate dehydrogenase, linked to at least one regulatory sequence for
XX transcription or translation in plants. The method is used to produce
XX chemicals, particularly Vitamins E or K and/or ubiquinone, useful e.g. as
XX fat-soluble antioxidants (in feeds, pharmaceuticals and cosmetics), but
XX also aromatic amino acids, salicylic or folic acid derivatives.
XX CC Phenylpropanoids, flavonoids etc., especially in transgenic plants. Also
XX CC plants with increased Vitamin E contents have improved resistance to
XX CC abiotic stress, e.g. frost and drought. Transgenic plants with an altered
XX CC shikimate pathway are useful as foods, fodder and in preparation of
XX CC processed foodstuffs. Transgenic plants with modified shikimate pathways
XX CC have increased content of desired chemicals, particularly tocopherols
XX CC and/or tocotrienols. This sequence encodes the Escherichia coli tyrA
XX CC (chorismate mutase prephenate dehydrogenase) protein used in the method
XX CC of the invention
XX
XX SQ Sequence 1238 BP; 269 A; 275 C; 370 G; 324 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 3,14e-05 Length: 1238
XX Score: 233.00 Matches: 77
XX Percent Similarity: 45.04% Conservative: 50
XX Best Local Similarity: 27.30% Mismatches: 111
XX Query Match: 14.25% Indels: 44
XX DB: Gaps: 8
XX
XX US-10-089-514-6 (1-322) x ABA99612 (1-1238)
XX
XX 8 ValValValGlyGlySerGlyAlaValAlaGlyMetPheAlaGlyLeuAlaArgGluAla 27
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 328 GTTATCGTCGCGCGGTGCGGTCAAGTGGACGCTGTTTCAGAGAAGATGACCTCTCG 387
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 28 GlySerArgThrLeuValValAlaLeuValProProProGlyArgProAspAlaCysLeu 47
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 388 GGTATATCAGGTGCGGATTCGAG-----CAACATGACTGGGATCGA 429
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 48 ValGlyAspValThrAlaProGlyProGluLeuAlaAlaLeuAlaAspAlaAspLeu 67
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 430 GCGGCTGATATTTGTTGCC-----GATGCCGGAATG 459
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 68 ValLeuLeuAlaValAlaHisGluAspValAlaLeuAlaValAlaProValThrArgLeu 87

```

```

DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 460 GTATTGTTAGTGTGCGCAATCCAGCTTACGACGAAGTATTGGCAATATACCGCCTTAA 519
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 88 MetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrGly---MetAlaAla 106
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 520 CCAGAA---GATTGTATCTGTGTCGATCTGCGCATGTGAGTAAGTAATGGCCATACGCGC 576
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 107 GluLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeuAsnProMetPheAlaPro 126
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 577 ATGCTGTGGCGCATATGATGTCGAGT-----CTGGGCTACACCGAGTTTCGCTCG 630
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 127 AlaAlaGlyMetThrGlyArgProValAlaAlaValAlaThrArgAspGlyProGlyVal 146
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 631 GACAGCGGTAGCCTGCGCAAGCAAGTT-----GTGCTGCTGATGACCTAAACCG 684
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 147 ThrAla-----LeuLeuArgLeuValGluGlyGlyValArgProValArgLeu 163
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 685 GAAGCATACCAATGTTCTCGAGCAAAATTCAGTCTGGGCGCTGCGCATCGTATT 744
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 164 ThrAlaGluGlnHisAspArgThrThrAlaAlaThrAlaAlaLeuThrHisAlaValLeu 183
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 745 AGCGCGTGCAGCAGCATGCAATATGCGCTTATTTCAGGCACTGCGCATTTGCTACT 804
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaAlaThr 203
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 805 TTTCCTTACGGGCTGCACTGCGCAAGAAATGTTACAGTTCAGCACTTCGCGCTC 864
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 204 AlaProProHisGlnValLeuLeuAlaLeuAlaArgValLeuGlySerPro 223
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 865 TCTTCGCCGATTATACCGCCTTGAGCTGCGCATGTCGCGGCACTGTTGCTCAGATCCG 924
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 224 GluValTyrGlyAspIleGlnArgSerAsnProArgAlaAlaSerAla----- 239
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 925 CACCTTATGCCAGCATCATATATGTGTGACAGCCGTAATCTGCGTTAAATCAACGTTAC 984
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 240 -----ArgArgAlaLeuAla 244
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 985 TATTAACGTTTCGCGCAGCGATTTGCTGCTGACAGCGCATTAAGACGCGTTTATT 1044
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 245 GluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArgAlaAspAlaPro 264
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1045 GACAGTTTCGCGCAAGGTGAGCAGCTGTCGCGGATTACGACAGCGTTTTCAGAGTGA 1104
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 265 GlyArg 266
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1105 AGCCGC 1110
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

## RESULT 15

AL46032 standard; DNA; 1238 BP.

AL46032;

11-JUL-2002 (first entry)

E coli TyrA gene.

Vitamin E; homogenisate; HG; homogenisate-1,2-dioxygenase; HGD;  
 KM maleylacetate-isomerase; MAI; fumarylacetate-hydrolase; FAH;  
 KM transgenic; nutrition; antioxidant; cardiovascular disease; cancer;  
 KM immune system; generalised age-related degeneration; animal feed;  
 KM meat quality; cosmetics; growth regulator herbicide; cardiac;  
 KM cytosolic; immunostimulant; enzyme; TyrA; chorismatmutase;  
 KM prephenate dehydrogenase; gene; ds.

Escherichia coli.

```

XX OS Escherichia coli.
XX
XX Key Location/Qualifiers
XX FT CDS 25..1146
XX FT /*tag= a
XX FT /product= "bifunctional chorismatmutase/
XX FT prephenate dehydrogenase"
XX

```







GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 10:07:08 ; Search time 176.216 Seconds  
(without alignments)  
2989.973 Million cell updates/sec

Title: US-10-089-514-6  
Perfect score: 1635  
Sequence: 1 MSGFPRSVVVGSGAVGWMF.....HCOELPRLHRTDDEKXDR 322

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapext 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 81313359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool\_h/US10089514/runat\_04102005\_105746\_8071/app\_query.fasta\_1.1621  
-DB=Issued Patents NA -OPMT=fastrap -SUFFIX=rnt -MINMATCH=0.1 -LOOPT=0  
-LOOEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blotsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10089514.@CGN\_1.116 @runat\_04102005\_105746\_8071 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	661.5	40.5	888	3 US-08-765-907A-2	Sequence 2, Appl1
2	661.5	40.5	888	4 US-09-987-614A-1	Sequence 2, Appl1
3	661.5	40.5	2888	4 US-08-765-907A-2	Sequence 1, Appl1
4	661.5	40.5	2888	4 US-09-987-614A-1	Sequence 1, Appl1
5	235	14.4	1664976	4 US-08-916-421B-1	Sequence 1, Appl1
6	235	14.4	1664976	4 US-09-692-570-1	Sequence 1, Appl1
7	227	13.9	1131	4 US-09-489-039A-4572	Sequence 4572, Ap
8	216	13.2	1830121	4 US-09-557-884-1	Sequence 1, Appl1
9	216	13.2	1830121	4 US-09-643-990A-1	Sequence 1, Appl1
10	200	12.2	1125	4 US-09-543-681A-1138	Sequence 1138, Ap
11	166.5	10.2	1038	4 US-09-902-540-5637	Sequence 5637, Ap
12	166.5	10.2	49225	4 US-09-902-540-1269	Sequence 1269, Ap

C	13	163.5	10.0	4411529	3	US-09-103-840A-1	Sequence 1, Appl1
C	14	150.5	9.2	1362	4	US-09-724-797-31	Sequence 31, Appl1
C	15	149	9.1	2732	4	US-09-589-892B-3	Sequence 3, Appl1
C	16	149	9.1	3810	4	US-09-902-540-3021	Sequence 3021, Ap
C	17	149	9.1	3812	4	US-09-902-540-641	Sequence 641, Ap
C	18	147	9.0	1848	4	US-09-902-540-3320	Sequence 3320, Ap
C	19	147	9.0	19302	4	US-09-902-540-1155	Sequence 1155, Ap
C	20	146.5	9.0	2376	4	US-09-252-991A-6630	Sequence 6630, Ap
C	21	146	8.9	2163	4	US-09-252-991A-5310	Sequence 5310, Ap
C	22	146	8.9	3282	4	US-09-252-991A-5256	Sequence 5256, Ap
C	23	146	8.9	4158	4	US-09-252-991A-5348	Sequence 5348, Ap
C	24	146	8.9	4953	4	US-09-252-991A-5227	Sequence 5227, Ap
C	25	145	8.9	3600	1	US-08-537-002A-5	Sequence 5, Appl1
C	26	145	8.9	3600	3	US-08-863-010-5	Sequence 5, Appl1
C	27	145	8.9	3600	3	US-09-024-429-5	Sequence 5, Appl1
C	28	145	8.9	3957	4	US-10-237-551-193	Sequence 193, Ap
C	29	145	8.9	154746	4	US-09-827-688-8	Sequence 8, Appl1
C	30	145	8.9	154746	4	US-09-827-688-8	Sequence 8, Appl1
C	31	144	8.8	4403765	3	US-09-103-840A-2	Sequence 2, Appl1
C	32	141.5	8.7	1527	4	US-09-248-796A-4088	Sequence 4088, Ap
C	33	140	8.6	21010	4	US-09-902-540-1188	Sequence 1188, Ap
C	34	140	8.6	4403765	3	US-09-103-840A-2	Sequence 2, Appl1
C	35	140	8.6	4411529	3	US-09-103-840A-1	Sequence 1, Appl1
C	36	139.5	8.5	1289	3	US-09-385-028-21	Sequence 21, Appl1
C	37	139.5	8.5	1289	4	US-09-726-614-21	Sequence 21, Appl1
C	38	139.5	8.5	1289	4	US-09-385-028-21	Sequence 21, Appl1
C	39	139.5	8.5	11604	3	US-09-385-028-13	Sequence 13, Appl1
C	40	139.5	8.5	11604	4	US-09-726-614-13	Sequence 13, Appl1
C	41	139.5	8.5	11604	4	US-09-385-040-13	Sequence 13, Appl1
C	42	139.5	8.5	15079	3	US-09-385-028-1	Sequence 1, Appl1
C	43	139.5	8.5	15079	4	US-09-726-614-1	Sequence 1, Appl1
C	44	139.5	8.5	15120	4	US-09-385-040-1	Sequence 1, Appl1
C	45	139	8.5	2175	4	US-09-252-991A-15144	Sequence 15144, A

## ALIGNMENTS

RESULT 1  
US-08-765-907A-2  
; Sequence 2, Application US/08765907A  
; Patent No. 6352839  
; GENERAL INFORMATION:  
; APPLICANT: BLANC, Veronique  
; APPLICANT: THIBAUT, Denis  
; APPLICANT: BEMAS-JACQUES, Nathalie  
; APPLICANT: BLANCHE, Francis  
; APPLICANT: COUZET, Joel  
; APPLICANT: BARRIERE, Jean-Claude  
; APPLICANT: DEBUSCHE, Laurent  
; APPLICANT: FAMECHON, Alain  
; APPLICANT: PARIS, Jean-Marc  
; APPLICANT: DUTRUC-ROSSET, Gilles  
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By  
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By  
; FILE REFERENCE: Streptogramin genes  
; CURRENT APPLICATION NUMBER: US/08/765, 907A  
; CURRENT FILING DATE: 1997-03-20  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 888  
; TYPE: DNA  
; ORGANISM: Streptomyces pristinaespiralis  
US-08-765-907A-2

## Alignment Scores:

Pred. No.: 4,81e-51  
Score: 661.50  
Percent Similarity: 60.38%  
Best Local Similarity: 49.52%  
Query Match: 40.46%  
DB: 3  
Length: 888  
Matches: 155  
Conservative: 34  
Mismatches: 101  
Indels: 23  
Gaps: 4



```
US-10-089-514-6 (1-322) x US-08-765-907A-2 (1-888)
QY      2 SerGlyPheProAArgSerValValValGlyGlySerGlyValValGlyGlyMetPheAla 21
      13 TCGGTGTTCCGGGCGTTGTGTGTGTGTGGGGCGGCGGTGCGGTGGCCGATGTTTCAGC 72
QY      22 GlyLeuLeuAArgGluAglYSerArgThrLeuValValAspLeuValProProProGly 41
      73 CACTGCTGTGCTGTGGGGGTGGCCGTGACCTGCTGAC---GTGGCCGGGGCGGCT 129
QY      42 ArgProAspAlaCysLeuVal-----GlyAspValThrAlaProGlyProGluLeuAla 59
      130 GCGCGGACCGGGCGCGGTGTGTGGCCGTGATGTGGCGCGCGCGGGGCGAGGCGCTC 189
QY      60 AlaAlaLeuAArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuLys 79
      190 GCGGCGCTGGCGGGCGGACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 249
QY      80 AlaValAlaProValThrArgLeuMetArgProGlyValAlaLeuLeuAlaAspThrLeuSer 99
      250 GCGGTGAGGTGTGTGGCGGGGTGTGATGTGGCCGCGGTGTGTGTGTGTGTGTGTGTGT 309
QY      100 ValArgThrGlyMetAlaAlaGluLeuAlaAlaHisAlaAlaProGlyValGlnHisValGly 119
      310 GTCAAGAGCGGATCGCCGGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 369
QY      120 LeuAsnProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValAla 139
      370 CTGAACCCGATGTTTCCGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 429
QY      140 ThrArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGluGlyGlyGlyArgX 159
      430 GTCAACCAACGGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 489
QY      160 ProValArgLeuThrArgGluGluGluHisAspArgThrThrAlaAlaThrGlnAlaLeuThr 179
      490 GTGTGTGAGATGCCGGCGCGCGCGGACGACGACGACGCGCGGACGACGCGCGCGCGC 549
QY      180 HisAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAla 199
      550 CATGCCCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 609
QY      200 LeuAlaAlaThrAlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeu 219
      610 CTCGGGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 669
QY      220 GlyGlySerProGluValThrGlyAspGlnArgSerAsnProArgAlaAlaSerAla 239
      670 GGGGAGGCGCGGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 729
QY      240 ArgArgAlaLeuAlaGluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAsp 259
      730 CGGACGGCGCTGTGGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 777
QY      260 ArgAlaAspAlaProGlyArgAlaAspAlaProGlyHisProGlyGlyCysAspGlyAla 279
      778 AGGGGCGACGACGAGACGTTCGCC----- 801
QY      280 GlyAsnLeuAspGlyValPheGlyGluLeuArgArgLeuMetGlyProGluLeuAlaAla 299
      802 -----GCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 849
QY      300 GlyGlnAspHisCysGlnGluLeuPheArgThrLeuHis 312
      850 CTGGAACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 888
RESULT 2
US-09-987-614A-2
; Sequence 2, Application US/09987614A
; Patent No. 6833382
; GENERAL INFORMATION:
; APPLICANT: BLANCO, Veronique
; APPLICANT: THIBAUT, Denis
```

```
; APPLICANT: BAMBAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSETT, Gilles
; TITLE OF INVENTION: Streptogramin And Method For Preparing Same By
; FILE REFERENCE: MutaSynthesis
; CURRENT APPLICATION NUMBER: US/09/987,614A
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US/08/765,907
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 888
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-2
```

## Alignment Scores:

Pred. No.:	4,81e-51	Length:	888
Score:	661.50	Matchee:	155
Percent Similarity:	60.38%	Conservative:	34
Best Local Similarity:	49.52%	Mismatches:	101
Query Match:	40.46%	Indels:	23
DB:	4	Gaps:	4

US-10-089-514-6 (1-322) x US-09-987-614A-2 (1-888)

```
QY      2 SerGlyPheProAArgSerValValValGlyGlySerGlyValValGlyGlyMetPheAla 21
      13 TCGGTGTTCCGGGCGTTGTGTGTGTGTGGGGCGGCGGTGCGGTGGCCGATGTTTCAGC 72
QY      22 GlyLeuLeuAArgGluAglYSerArgThrLeuValValAspLeuValProProProGly 41
      73 CACTGCTGTGCTGTGGGGGTGGCCGTGACCTGCTGAC---GTGGCCGGGGCGGCT 129
QY      42 ArgProAspAlaCysLeuVal-----GlyAspValThrAlaProGlyProGluLeuAla 59
      130 GCGGCGCTGGCGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 189
QY      60 AlaAlaLeuAArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuLys 79
      190 GCGGCGCTGGCGGGCGGACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 249
QY      80 AlaValAlaProValThrArgLeuMetArgProGlyValAlaLeuLeuAlaAspThrLeuSer 99
      250 GGGTGTGAGATGCCGGCGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 309
QY      220 GlyGlySerProGluValThrGlyAspGlnArgSerAsnProArgAlaAlaSerAla 239
      250 GGGTGTGAGATGCCGGCGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 309
QY      100 ValArgThrGlyMetAlaAlaGluLeuAlaAlaHisAlaAlaProGlyValGlnHisValGly 119
      310 GTCAAGAGCGGATGCCCGGGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 369
QY      120 LeuAsnProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValAla 139
      370 CTGAACCCGATGTTTCCGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 429
QY      140 ThrArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGluGlyGlyGlyArgX 159
      430 GTCAACCAACGGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 489
QY      160 ProValArgLeuThrArgGluGluGluHisAspArgThrThrAlaAlaThrGlnAlaLeuThr 179
      490 GTGTGTGAGATGCCGGCGCGGACGACGACGACGACGCGCGGACGACGCGCGCGCGC 549
QY      180 HisAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAla 199
      550 CATGCCCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 609
```



```
QY      200  LeuAAlaAlaThlArProPProHfsgInValLeuAlaLeuAlaArgValLeu 219
Db      610  CTGCGGAGACATGTCCGCCGCCGCAGATTCTGGCATGTGGCTGTGGTGGCCGGAATGCC 669
QY      220  GtGlySerProGluValTyrgIAspRIegInArgSerxanProARgAlaAlaSerAla 239
Db      670  GCGCGGGACGCCCGGAGGTGTAATTTCGACATTCGACGCCGCAACC CGCGCGCGCGCGCG 729
QY      240  ArgArGAlaLeuAlaGluAlaLeuArgSerPheAlaAlaLeuValGtIAspAProAsp 259
Db      720  CCGCAGGCGCTGGGCGCGCGCGCTG-----CTGCGCGTGGGGCAGCGCGCTGCAG 777
QY      260  ArgAlaAerAlaRProGIyArGAlaAerAlaRProGIyNIeRProGIyCyAspRIyAla 279
Db      778  AGGGCGCGAGCAGGAGACGTTCCG----- 801
QY      280  GluAenLeuAerPluYalPheGtGlUleuArgArgLeuMetGlyProGluLeuAlaAla 299
Db      802  -----GCCCTGTTCGCGCAACTGCGGGTGTCGTGGCGAGCACCGCGCGAG 849
QY      300  GluGlnAerPIeCyGtIngluleuPheArGTrhLeuNI s 312
Db      850  CTGGAACGCGTGTGCGCGCGAGATGTTCAACGCCCTGCAC 888
```

### RESULT 3

```

US-08-765-907A-1/C
: Sequence 1, Application US/08765907A
: Patent No. 6352839
:
: GENERAL INFORMATION:
: APPLICANT: BLANC, Veronique
: APPLICANT: THIBAUT, Denis
: APPLICANT: BAMA-JACQUES, Nathalie
: APPLICANT: BLANCHE, Francis
: APPLICANT: COUZET, Joel
: APPLICANT: BARRIERE, Jean-Claude
: APPLICANT: DEBUSCHÉ, Laurent
: APPLICANT: FAMECHON, Alain
: APPLICANT: PARIS, Jean-Marc
: APPLICANT: DUTRUC-ROSSET, Gilles
: TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
: TITLE OF INVENTION: Mutasynthesis
: FILE REFERENCE: Streptogramin genes
: CURRENT APPLICATION NUMBER: US/08/765,907A
: CURRENT FILING DATE: 1997-03-20
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 2888
: TYPE: DNA
: ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-1

```

**Alignment Scores:**

Pred. No.:	2,48e-50	length:	2888
Score:	661.50	Matches:	155
Percent Similarity:	60.38%	Conservative:	34
Best Local Similarity:	49.52%	Mismatches:	101
Query Match:	40.46%	Indels:	23
DB:	3	Gaps:	4

US-10-089-514-6 (1-322) X US-08-765-907A-1 (1-2888)

QY	2	SergIyGlyPheProArgSerValValValGlyGlySerGlyValaValGlyMetPheAla	21
Db	1824	TCGGTTCGCGCGCTGTGTGTGTGTGTGGCGCGCGCGCGGTGGCGCATGTTCAGC	1765
QY	22	GlyLeuLeuArgGluAlaGlySerArgThrLeuValValaPheLeuValProProProGly	41
Db	1764	CAC TGG CCG TGG TCG TTT CGG GGG GTG GCG GGT GAC CTG GCG GAC -- GTG ACC GGG GCG CGT	1708
QY	42	ArgProAlaAlaCysLeuVal-----GlyAspValThrAlaProGlyProGlyIleuAla	59
Db	1707	GCGCGCGACCGGGGTGCGGGTGGTGTGCGCCCGGTATGTGTGCGCGCGCGCCCGGAGCGCGTGC	1648

OY		60	AlaIleuArgAspAlaAspLeuValIleuLeuAlaHisGluAspValAlaIleuLys	79
Dd		1647	GCGGCCTGGCGGCGCGAGCTGTGTGTCTGGCCGATCCCGAGCCGGTGGCGGAG	1588
OY		80	AAlaValAProValThrArgLeuMetArgProGlyAlaIleuLeuAlaAspThrLeuSer	99
Dd		1587	GCGGTGAAGGTGTGGCGGGGTGATGCGGCCGCTCGGGTGCTCGGGACACTTGTCG	1528
OY		100	VAlaArgThrGlyMetAlaGluLeuAlaHisAlaProGlyValGlnHisValGly	119
Dd		1527	GTCAAAGACCAGAATCGCGGGCGGCTCTGTGAAGCGCGCGCGGGCTGCAGGCCGTGGGG	1468
OY		120	LeuAsnProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValAl	139
Dd		1467	CTGAACCGAATTGTGCCCCCTCTCTGTGGTCTTTAGAGGGCGGCCGTGGCGGTGTGTG	1408
OY		140	ThirArgSpgIyProGlyValThrAlaLeuLeuArgLeuValGluGlyGlyValArg	159
Dd		1407	GTCAACGACGGCGCGGGTGTGGCGGCTCTGTGTGAGCTGGTGGCGGGTGGGGGGCCGG	1348
OY		160	ProValArgLeuThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThr	179
Dd		1347	GTCGTGAGAGATCGCGGCGCGCGCACAAGAGACTGACCCGCGCGCACAGCGCCACG	1288
OY		180	HisAlaValLeuLeuSerPheGlyLeuAlaIleuAlaArgLeuGlyValAspValArgAla	199
Dd		1287	CATGCGCGGTGTCTGGCTTCTGGAGCTGGCGCTGGGTAGCTGTGTGTGACGTGGGGCG	1228
OY		200	LeuAlaAlaThrAlaProProPheHisGlnValLeuLeuAlaLeuLeuAlaArgValLeu	219
Dd		1227	CTGGGGGACAGTGCCTCCGCGCGCCCATCTGGCCAGTGTGGGCTGTGGCCGGATCGCC	1168
OY		220	GlyIySerProGluValTyrgLyAspTleglInArgSerAsnProArgAlaAlaSerAla	239
Dd		1167	GCGCGGACGCGCGAGGTGTATTTTCAGATCAGGCGCCCAACCCGCGCGCGCGCGCG	1108
OY		240	ArgArgAlaIleuAlaGluAlaLeuArgSerPheAlaIleuValIglYAspAspProAsp	259
Dd		1107	CGGACGGGGCTGGGCGCGCGCTCT-----GTGCGGCTGGGGACAGCCGTGGAG	1060
OY		260	ArgAlaAspAlaProGlyArgAlaAspAlaProGlyHisProGlyGlyCyAspGlyAla	279
Dd		1059	AGGGGCGACGAGAGACGTTGCC-----	1036
OY		280	GlyAsnLeuAspGlyValPheGlyGlnLeuArgArgLeuMetGlyProGluLeuAlaAla	299
Dd		1035	-----GCCCTGTTCGCCGAACCTGGCGCGTGTGTGGCGGACGCGCGCGAG	988
OY		300	GlyGlnAspHisCysGlnGluLeuPheArgThrLeuHis	312
Dd		987	CTGAAACCGCTGTGCGCGGAGATTCAACGCCCTGTGAC	949

## RESULT 4

```

US-09-987-614A-1/C
Sequence 1, Application US/09987614A
Patent No. 683382
GENERAL INFORMATION:
APPLICANT: BLANC, Veronique
APPLICANT: THIBAUT, Denis
APPLICANT: BAMA-JACQUES, Nathalie
APPLICANT: BLANCHE, Francis
APPLICANT: COUZET, Joel
APPLICANT: BARRIERE, Jean-Claude
APPLICANT: DEBUSCHÉ, Laurent
APPLICANT: FAMECHON, Alain
APPLICANT: PARI, Jean-Marc
APPLICANT: DUTRUC-ROSSET, Gilles
TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
TITLE OR INVENTION: Metasynthesis
FILE REFERENCE: Streptogramin genes
CURRENT APPLICATION NUMBER: US/09/987,614A
CURRENT FILING DATE: 2001-11-15

```







```
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (191995) .. (191995)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (231980) .. (231980)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (234187) .. (234187)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (234220) .. (234220)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (234814) .. (234814)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (309398) .. (309398)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (309418) .. (309418)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (312837) .. (312837)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (312993) .. (312993)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (319226) .. (319226)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (559167) .. (559167)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (559241) .. (559241)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (600992) .. (600992)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (622708) .. (622708)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (657081) .. (657081)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (657203) .. (657203)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (674435) .. (674435)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (682442) .. (682442)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (713652) .. (713652)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (741684) .. (741684)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (779455) .. (779455)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (779676) .. (779676)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (855539) .. (855539)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (871619) .. (871619)
? OTHER INFORMATION: n equals a, t, c, or g
```

```
? NAME/KEY: misc feature
? LOCATION: (1084830) .. (1084830)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (1096846) .. (1096846)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (1119881) .. (1119881)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (1130881) .. (1130881)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (1310988) .. (1310988)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (1313224) .. (1313224)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (1349473) .. (1349473)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (1349491) .. (1349491)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (1470091) .. (1470091)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (1569020) .. (1569020)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (1602912) .. (1602912)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (1603734) .. (1603734)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (1637998) .. (1637998)
? OTHER INFORMATION: n equals a, t, c, or g
? NAME/KEY: misc feature
? LOCATION: (1664854) .. (1664854)
? OTHER INFORMATION: n equals a, t, c, or g
? US-08-916-421B-1

Alignment Scores:
Pred. No.: 9,71e-08 Length: 1664976
Score: 235.00 Matches: 71
Percent Similarity: 42.76 Conservative: 59
Best Local Similarity: 23.36 Mismatches: 158
Query Match: 14.378 Indels: 16
DB: Gaps: 5

US-10-089-514-6 (1-322) x US-08-916-421B-1 (1-1664976)
QY 9 ValValGlyGlySerGlyValValGlyGlyMetPheAlaGlyLeuLeuArgGluAlaGly 28
DB 930033 ATTATTGGAGGAACTGATGCTTTAGGGAAGTGGTTGCAAGATTATTAATAAAGG 929974
QY 29 SerArgThrLeuValValValLeuValProProGlyArgProAspAlaCysLeuVal 48
DB 929973 TTTAACGTTATAGCTTACTGCGAGAGATATTGGAAGAAAGAAAATCTTGAAAGCTTA 929914
QY 49 GlyAspValThrAlaProGlyProGluLeuAlaAlaLeuArgAspAlaAspLeuVal 68
DB 929913 GGG---GTTGAATTTACTAACAACAACATTTGAAGCTGCT---AAAAAGGAGATATTGTT 929860
QY 69 LeuLeuAlaValHisGluAspValAlaLeuValAlaValAlaProValThrArgLeuMet 88
DB 929859 ATTGTAGCAGTTCCAAATAATGTTACAGAAAGGTTTAAAGAGTGCTCTCATGTT 929800
QY 89 ArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrGlyMetAlaAlaGluLeu 108
DB 929799 AGGGAAGGATGTTTATTATATGACATACCTCAATTAAAGAGATTCTTCAAGAGCTATG 929740
```







	OTHER INFORMATION:	n	equals	a,	t,	c,	or	g
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(309418) .. (309418)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(312637) .. (312637)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(312993) .. (312993)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(319226) .. (319226)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(559167) .. (559167)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(559241) .. (559241)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(600992) .. (600992)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(622708) .. (622708)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(657081) .. (657081)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(657203) .. (657203)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(674435) .. (674435)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(682442) .. (682442)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(713652) .. (713652)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(741684) .. (741684)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(779455) .. (779455)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(779676) .. (779676)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(855539) .. (855539)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(871619) .. (871619)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	

[illegible]



```

Db 929559 CCTCCAGAAAAACATGATAGAAATTATGGGATTTGTTCAAGGTTTGACTCACTTGCCCTTT 9295600
Qy 184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaIaThr 203
Db 929499 ATATCTTAGAGCAAGCTTAAAGAACTCAACGTTTATATATAAGAGTCAAGAAAGTTT 929440
Qy 204 AlaProProProHisGlnValLeuLeuAlaLeuAlaArgValLeuGlyGlySerPro 223
Db 929439 GCTCCCCCAATATACGATGATGATTTCTATCATTTGGAGAGATTAAGAGCAAGAAATCCC 929380
Qy 224 GluValIYrGlyAspIleGlnArgSerAspArgAlaIaIaSerAlaArgAlaLeu 243
Db 929379 TATTTATATCTGACATCCAAATGTTTAAATCAAGGTAAGAAGATTCATGAAGCTTTT 929320
Qy 244 AlaGluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArgAlaAspAla 263
Db 929319 ATAAATAGTGTAGAGAAATACAGTGAATTTGTTAAAAATAAAGATAGAGAGTTTGT 929260
Qy 264 ProGlyArgAlaAspAlaProGlyHisProGlyGlyCysAspGlyAlaGly----- 280
Db 929259 AAGATATGAAGAAGACTGCAAGCAATTTTGTAGTGAGCAAAAGAGAGGCGCTTATAT 929200
Qy 281 -----AenLeuAspGlyValPheGlyGluLeuArgAlaGluMetGly 294
Db 929199 TCAGATAGAGCCGCTGTTGCAATTAAACATCTGAAATGGAAGCTAAATAAATCAATTGT 929140
Qy 295 ProGluLeuAla 298
Db 929139 AAAGATGTGCT 929128

RESULT 7
US-09-489-039A-4572
; Sequence 4572, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 4572
; LENGTH: 1131
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-4572

Alignment Scores:
Pred. No.: 2,03e-11 Length: 1131
Score: 227.00 Matches: 83
Percent Similarity: 45.80% Conservative: 48
Best Local Similarity: 29.02% Mismatches: 107
Query Match: 13.88% Indels: 49
DB: 4 Gaps: 9

US-10-089-514-6 (1-322) x US-09-489-039A-4572 (1-1131)
Qy 8 ValValValGlyGlySerGlyAlaValGlyGlyMetPheAlaGlyLeuLeuArgGluAla 27
Db 313 GTATGCTGCGCGCGCGCGGAGAGATGGCGCGCTGTTGAGAAAGATGTAACGCTGTC 372
Qy 28 GlySerArgThrLeuValAlaAspLeuValProProProGlyArgProAlaCysLeu 47
Db 373 GGCCTACCAAGGTGGCTATCTGAA-----AAAAACGATTGGGCCCGG 414
Qy 48 ValGlyAspValThrAlaProGlyProGluLeuAlaAlaIaLeuArgAspAlaAspLeu 67
Db 415 GCGCGGATATCGTCGC-----GATGCGCGGATG 444
Qy 68 ValLeuLeuAlaValHisGluAspValAlaLeuAlaValAlaProValThrArgLeu 87

```

```

Db 445 GTTATGTCAGCGCTACCGATTCAACACACCGGTGAGACAGATCGGACGCTGCGCCCTT 504
Qy 88 MetArgProGlyValaLeuLeuAlaAspThrLeuSerValArgThr---GlyMetAlaAla 106
Db 505 ---CCGGCGGATTCATTCGTGTTGACCTTGCTCGGTGTAAGCGGACCGCGCGGG 561
Qy 107 GluLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeuAsnProMetPheAlaPro 126
Db 562 ATGCTGCGCGCGATCAGGCGCCCGTT-----CTTGCGCTGACCCGATGTTCCGTCCG 615
Qy 127 AlaAlaGlyMetThrGlyArgProValAlaAlaValAlaThrArgAspGlyProGlyVal 146
Db 616 GACAGCGGACGCGCTGCGCAAGCAGGTG-----GTGCTATATGTTATGTCGCGCCAGCGG 669
Qy 147 ThrAla-----LeuLeuArgLeuValGluGlyGlyGlyValArgProValArgLeu 163
Db 670 GAGGCTTATCAGTGTCTCTCGAACAATCCAGTGTGGGGCGCGCGCTGCACCGCATC 729
Qy 164 ThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeu 183
Db 730 AGGCGCGTTGAGACGATCAGAACATGCGCTTATTCAGGCGCTGCGCATTTGCGCAC 789
Qy 184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaIaThr 203
Db 790 TTCGCTATGCGCTGCACTGCGGAGAAATGTACGTTCAGACCACTGCTGCGGCTC 849
Qy 204 AlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGlySerPro 223
Db 850 TCGTCAACCATTTTACCGACTGAGCTGGCGAGTGTGGCGCTGTTCCCATGATCCG 909
Qy 224 GluValIYrGlyAspIleGlnArgSerAspProArgAlaAlaSerAlaArgAlaLeu 243
Db 910 CAGCTTAAGCGGATTTTATATGTCATCGGA-AAACAACCGGCGCTATTAAAGCGTA 968
Qy 244 AlaGluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArgAlaAspAla 263
Db 969 CTACCAAGCTTTCCG-----GAGCAGCGCGATAGCA 1019
Qy 264 ProGlyArgAlaAspAlaProGlyHisProGlyGlyCysAspGlyValaGlyAsnLeuAsp 283
Db 984 -----CGAAGCATGCGCTGCT-----GAGCAGCGCGATAGCA 1019
Qy 284 GlyValPheGlyGluLeu 289
Db 1020 GCGCTTATCGACACGTT 1037

RESULT 8
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>

```











```

; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 5637
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-5637

Alignment Scores:
Pred. No.: 5,64e-06 Length: 1038
Score: 166.50 Matches: 86
Percent Similarity: 41.22% Conservative: 36
Best Local Similarity: 29.05% Mismatches: 144
Query Match: 10.18% Indels: 31
DB: 4 Gaps: 10

US-10-089-514-6 (1-322) x US-09-902-540-5637 (1-1038)
OY 7 ServaIvalValGlySerGlyValAlaValGlyMecPheAlaGlyLeuAlaArgGlu 26
DB 10 AGCATGCGCGTGTGGGATACGCGCGCTTGGCGCTGCTGCTGCTGCTGCTGCTG 69
OY 27 AlAGlySerArgThrLeuValAlaAspLeuValProProGlyIArg-----Pro 43
DB 70 GCGGCG-----ATCCCGCATCGGCTGTTCGAGCCGCGCAGATGACGTCGCC 117
OY 44 AspAlaCysLeuValGlyAspValThrAlaProGlyProGluLeuAlaAlaLeuArg 63
DB 118 GACGCA-----CTGAGGGCACCAGCGCTCGCGGAGAGTGTGAG 156
OY 64 AspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuLysAlaAlaPro 83
DB 157 GCGCGCGGCTGTGGTGTCTTCATGCGCGTCTGGAATGCGCTCGCTGAGAG 216
OY 84 ValThrArgLeuMecArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrGly 103
DB 217 CTTCGCGCGCGCTGTGCGCGCGCAGACGCTGCTGACCTCGGAAGCTCAAGTGC 276
OY 104 MetAlaAlaGluLeuAlaAla--HisAlaProGlyValGlnHisValGlyLeuAspPro 122
DB 277 CCTGTGACAGTGTGCGCTCTCGGTGGGCGGAAATTCCTCGGTGGGACAGCATCCG 336
OY 123 MetPheAlaProAlaAlaGlyMetThrArgProValAla-----AlaValAla 139
DB 337 CTGTTCGCGCGCGCAGTCT--GGCGAGGGCGACCTGCTCGCGCAGCGTGTGTGCC 395
OY 140 ThrArgAspGlyPro--Gly--ValThrAlaLeuLeuArgLeuValGlyGlyGly 158
DB 396 GAACGAACTGCACCCAGAGCGCTGCGCAGAGCCCGGCTCTTTCGAAGCATTCGGCTG 455
OY 158 ValArgProValArgLeuThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLe 178
DB 456 TGAAGTGCAGGAGCTGTGCGCGGATGCGACGACGCGCTGATGTGCGCAGCATGTGCT 515
OY 178 ValThrAlaAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAlaValAla 198
DB 516 CACC-----TCTTCTCGCGCCACGAGGTGTCTCAAGCGGAGCGCGGAAAGACCTG-- 567
OY 198 GalAlaLeuAlaAlaThrAlaProProHisGlnValLeuLeuAlaLeuLeuAlaArgVal 218
DB 568 -----CCTTTCGCGCGCGCCAGCTTCAGCCCGCTGCTGCTGAGAGATGACG 617
OY 218 IleuGlyGlySerProGluValIArgAspIleGlnArgSerAsnProArgAlaAlaLe 238
DB 618 ACCGCTGAGAGTGTCCATCTTTCGCGCGTGTGCGAGTCCGGAATCCCTCAACCGCGGGA 677
OY 238 rAlaArgArgAlaLeuAlaGluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspPr 258
```

```

DB 678 CGACCGCGCGCGCTGTGAGAGCCCTTCAAGCAGATCCAC-----CTGGCTGTGATTC 721
OY 258 oAspArgAlaAspAlaProGlyIArgAlaAspAlaProGlyHisProGlyIArgAspArg 278
DB 732 GCGCGCGCTGTGGGCGCGCGGAGCTGTGCGCGCGCG-----CCGCGCTGACGGAAGT 785
OY 278 ValAGlyLeuLeuAspGlyValPheGlyLeuLeuArgLeuMet 293
DB 786 TCGGAGTGTGTGATGCTGTGACCGGTGACCGTGTGATTCCTG 831

RESULT 12
US-09-902-540-1269/c
; Sequence 1269, Application US/09902540
; Patent No. 683347
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1269
; LENGTH: 49225
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1269

Alignment Scores:
Pred. No.: 0.00121 Length: 49225
Score: 166.50 Matches: 86
Percent Similarity: 41.22% Conservative: 36
Best Local Similarity: 29.05% Mismatches: 144
Query Match: 10.18% Indels: 31
DB: 4 Gaps: 10

US-10-089-514-6 (1-322) x US-09-902-540-1269 (1-49225)
OY 7 ServaIvalValGlySerGlyValAlaValGlyMecPheAlaGlyLeuAlaArgGlu 26
DB 41669 AGCATGCGCGTGTGGGATACGCGCGCTTGGCGCTGCTGCTGCTGCTGCTG 41610
OY 27 AlAGlySerArgThrLeuValAlaAspLeuValProProGlyIArg-----Pro 43
DB 41609 GCGGCG-----ATCCCGCATCGGCTGTTCGAGCCGCGCAGATGACGTCGCC 41562
OY 44 AspAlaCysLeuValGlyAspValThrAlaProGlyProGluLeuAlaAlaLeuArg 63
DB 41561 GACGCA-----CTGAGGGCACCAGCGCTCGCGGAGAGTGTGAG 41523
OY 64 AspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuLysAlaAlaPro 83
DB 41522 GCGCGCGGCTGTGGTGTCTTCATGCGCGCTGTGGAATGCGCTCGCTGAGAGAG 41463
OY 84 ValThrArgLeuMecArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrGly 103
DB 41462 CTTCGCGCGCGCTGTGCGCGCGCAGACGCTGCTGGAATGCGCTCGCTGAGAGAG 41403
OY 104 MetAlaAlaGluLeuAlaAla--HisAlaProGlyValGlnHisValGlyLeuAspPro 122
DB 41402 CTGTGACAGTGTGCTGCGCGCGGAGCATTCCTCGGTGGGACAGCATTCG 41343
OY 123 MetPheAlaProAlaAlaGlyMetThrArgIArgProValAla-----AlaValAla 139
DB 41342 CTGTTCGCGCGCGCAGTCT--GGCGAGGGCGACCTGCTGCGCAGCGTGTGTGCC 41284
OY 140 ThrArgAspGlyPro--Gly--ValThrAlaLeuLeuArgLeuValGlyGlyGly 158
```







```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 1362
; TYPE: DNA
; ORGANISM: Bacteria
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1362)
US-09-724-797-31

Alignment Scores:
Pred. No.: 0.000234 Length: 1362
Score: 150.50 Matches: 101
Percent Similarity: 35.03% Conservative: 30
Best Local Similarity: 27.01% Mismatches: 136
Query Match: 9.20% Indels: 108
DB: 4 Gaps: 18

US-10-089-514-6 (1-322) x US-09-724-797-31 (1-1362)
QY 5 ProArSeryValValAGlyGlySerGlyAlaValAGlyMetPheAlaGlyLeuLeu 24
DB 34 CCCTGAGAGGTGGTCACTCCGGGCGCCGCTACGCTCGCTCTCGCCGTCACCTC 93
QY 25 ArgGluAlaGlySerArGThrLeuValValAspLeuValProProGlyArgProAsp 44
DB 94 GCCGCGCGCGCGCGCGAGGTGGTGGCGGTCGAC-----AGCGACCGCGGACCGTCGCG 147
QY 45 AlaCysLeuValAGlyAspValThrAlaProGlyProGlyLeuAlaAlaLeuArgAsp 64
DB 148 GACCTCGCGCGCGCGCGCGGTCCCGGCTCCCGAGCCCGCGCGCGCGCGCGCGAGC 207
QY 65 -----AlaAsp 66
DB 208 CTCGCGCGCGACCGGACGCTGACGCGCGACGCTGACGCGCGGTCCGCGCGCGAGC 267
QY 67 LeuValLeuLeuAlaVal-----HisGluAspValAlaLeuLys 79
DB 268 GTGTGTATCTGTGACGCTGCGGACCGCGACCGCGCGCGACGATGTCACCGAGCAG 327
QY 80 AlaValAlaProValThrArgLeu-----MetArgProGlyAlaLeuLeuAlaAsp 96
DB 328 CTGTGCGCGCGCGCGCGAGCATCGCCCGCGGTGGCGCGCGCGCGCACTGTGATCTC 387
QY 97 ThrLeuSerValArgThrGlyMetAlaAlaGluLeuAlaAla-----HisAlaPro 113
DB 388 AAGTCACGGGTCTCCCGCGGACACCGCGGACCGTCTGCGCCCTGCTGAGAGCGGC 447
QY 114 GlyValGlnHis-----ValGlyLeuAsnProMetPheAlaProAlaAlaGlyMet 130
DB 448 GGGCTGTGTGACGACGCGCGCATTCGGGGCTG-----GCCCTTGTGCGCGAGCGGCTCGCC 501
QY 131 ThrGlyArgProValAlaAlaValAlaValThrArgAspGlyProGlyValThrAlaLeuLeu 150
DB 502 GAGGGGGTGGCGCTGGCGCGCAGTGGCGAGC-----CTG 534
QY 151 ArgLeuValGluGlyGlyGlyArgProValArgLeuThrAlaGluGluHisAspArg 170
DB 535 CCGGTGTGTGTGGGTGGTGGCGGC-----CCGCGC 564
QY 171 ThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeuLeuSerPheGlyLeuAlaLeu 190
DB 565 AGCGCGCGCGCGCGCGACGATCTGTGCGGTCCGCG----- 600
QY 191 AlaArgLeuGlyValAspValArgAlaLeuAlaAlaThrAlaProProProHisGlnVal 210
DB 601 -----CTCGCGCTGCGCGCTCCGCGCAGTCCGCTCGCGCGCAATCGCGCGAGTGGTA-- 652
QY 210 IleuLeuAlaLeuLeuAlaArgValLeuGlyGlySerProGlyValArgLysPheGly 230
DB 653 -----AGCTCGCGACCACTGGTGGATGACGCGAAGC--GCGCATCGCA 697
QY 230 nArgSerAsnPro-----ArgAlaAlaSerAlaArgArgAlaLeuAlaGluAlaLeuAr 248
```

```

DB 698 ACGAACTCCGCCCGGTACTGTGCGGCGTCTGGGGGTGAGACGTCCTCGACGATCGGGC-- 755
QY 248 gSerPheAlaAlaLeuValAGlyAspAspProAspAlaAspAlaProGlyArgAlaAs 268
DB 756 -----GGCGAACACCTTCGCCAAGGGCAGACACATGTGTAACTTCG 796
QY 268 pAlaProGlyHisAspProGlyGlyCys-----AspGlyAlaGlyAsnLeuAspGly 284
DB 797 TGCTGCGCGG-----GATGGGTGTGCGGCGCTCTGCTGCTGACGAAAGACCCGTGAATGG 850
QY 284 yValPheGlyGlyLeuLeuArgArgLeuMetGlyProGlyLeuAla----- 298
DB 851 CGTGGCGGAGACG-----CCGGGACCGGGGCGGTGCCCTCGCGACGCTGAGAG 898
QY 299 -----AlaGlyGlnAspHisCysGlnGluLeuPheArgThrLeuHisArg----- 313
DB 899 CGGCGCGCGCGGTCAACGACGACATGCCCGCGCACACCCCGCGCTCATCGCCGACGAGC 958
QY 314 -----ThrAspAspGluGlyGlyLysAspArg 322
DB 959 TGGTCAAGCTGGGACGGGATCGGACGACGACGACGATCCG 998

RESULT 15
US-09-589-892B-3/C
; Sequence 3, Application US/09589892B
; Patent No. 6689583
; GENERAL INFORMATION:
; APPLICANT: Jennewein, Thomas
; APPLICANT: Laible, Goetz
; APPLICANT: O'Carroll, Donal
; APPLICANT: Eisenhaber, Frank
; APPLICANT: Rea, Stephen
; TITLE OF INVENTION: Chromatin-Regulator Genes
; FILE REFERENCE: 0652.1670001
; CURRENT APPLICATION NUMBER: US/09/589, 892B
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: US 08/945, 988
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: PCT/EP96/01818
; PRIOR FILING DATE: 1996-05-02
; PRIOR APPLICATION NUMBER: DE 195 16 776.7
; PRIOR FILING DATE: 1995-05-10
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 2732
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5' UTR
; LOCATION: (1)..(44)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (45)..(1283)
; FEATURE:
; NAME/KEY: 3' UTR
; LOCATION: (1284)..(2732)
US-09-589-892B-3

Alignment Scores:
Pred. No.: 0.000844 Length: 2732
Score: 149.00 Matches: 82
Percent Similarity: 44.96% Conservative: 34
Best Local Similarity: 31.78% Mismatches: 97
Query Match: 9.11% Indels: 45
DB: 4 Gaps: 14

US-10-089-514-6 (1-322) x US-09-589-892B-3 (1-2732)
QY 9 ValValAGlyGlySerGlyAlaValAGlyGlyMetPheAlaGlyLeuLeuArgGluAlaGly 28
DB 989 ATGTATGGCGCGCATCCACGCGTGTACACGTCCTCACGTAGTCCAGGTCAAGAGCGTAGCT 930
```



QY 29 SerAGThrLeuValValAspLeuValProProProGlyArgProAspAlaCysLeuVal 48  
 |||||  
 Db 929 GGGCGCCGTGACGGTGTGATCTGGCCCCGCGC-----CTCTGCTCTTC 885  
 QY 49 GlyAspValThrAlaProGlyProGluLeuAlaAlaLeuArgAspAlaAspLeuVal 68  
 ||||  
 Db 884 -----TGAGGTATGATCTCTCCACGGTACCTCATGACGAA----- 849  
 QY 69 LeuLeuAlaValHisGluAspValAlaLeuLysAlaValAlaProValThrArgLeuMet 88  
 |||||  
 Db 848 -----GCTGTTCTTGGGATCTTCTCCAGGGGTGGCGAGCCCCACCGCCCATATC 795  
 QY 89 ArgPro-----GlyAlaLeuLeuAlaAspThrLeuSerValArgThrGlyMetAlaAla 106  
 |||||  
 Db 794 CGTCGGAGAAATGACAGAGGTATATCGADATCCCTTCTGACACACG---ATTGGGCA 738  
 QY 107 GluLeuAlaAlaHisAlaAlaProGlyValGlnHisValGlyLeuAsnProMetPheAlaPro 126  
 |||||  
 Db 737 GTCATAGCCGACGGCGACGGGAGTTGCACTGTGATGGCGACGCCGGCTTGAAAGCG 678  
 QY 127 -----AlaAlaGlyMetThrGly-----ArgProValAlaAlaValAla 139  
 |||||  
 Db 677 CACCTGGCCCTGTGATTTAGGCAAACTTGTGCATGAGACGCCCGCGGACGAGCTTC 618  
 QY 140 ThrArgAspGlyProGlyValThrAlaLeuLeu-----ArgLeuValGluGlyGly 156  
 |||||  
 Db 617 AGTGGGTGCCACAGACAGCTCTGTGCACCTGCAGCCACAGCACCTGTGTGGGTGTAT 558  
 QY 157 GlyGlyArgProValArgLeuThr-----AlaGluGluHisAspArgThrThr 172  
 |||||  
 Db 557 GCC-----CTCACCAACACAGGTACTCATTTGATGTATACACGAAGCCCCGGG 513  
 QY 173 AlaAlaThrGlnAlaLeuThrHisAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArg 192  
 |||||  
 Db 512 AGGGCGCTCCAG-----GTCCACCTCATCTCTCAAGTATGATCGTCCAGATGCTGCG 459  
 QY 193 LeuGlyValAspValArgAlaLeuAlaAlaThrAlaProProProHisGlnValLeu 212  
 |||||  
 Db 458 CTGGGATTTAGCTC---CTGCTCCACGAGACGAGCGCCGCT-----CTGCTT 411  
 QY 213 AlaLeuLeuAlaArgValLeuGlyGlySer---ProGluValTyrGlyAspIle-GlnArg 231  
 |||||  
 Db 410 GGCTTTCTTCACACAGGTAGTTGGCCAAAGCTGGGTCCAGGTGCGGGGGGTCTTTGACCG 351  
 QY 231 GSerAsnProAlaArgAlaAlaSerAlaArgArgAlaLeuAlaGluAlaLeuArg 248  
 |||||  
 Db 350 GTGGTCCGCGCGAGACGCTCCCTTTCTTAAGTCTTTGTGGAACTGCTTGAAG 299



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 5, 2005, 09:57:48 ; Search time 708.632 Seconds  
(without alignments)  
3137.453 Million cell updates/sec

Title: US-10-089-514-6

Perfect score: 1635  
Sequence: 1 MSGFPRSVVVGSGAVGMF.....HCOELPRTLHRTDDEGKDR 322

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 7442561 seqs, 3452328358 residues

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODE=frame+ p2n.model -DEV=X1h  
-Q=/cgn2\_1/USPTO.spool.h/US10089514/runat\_04102005.105745.8055/app.query.fasta\_1.1621  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=trmb -MINMATCH=0.1  
-LOOPCU=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blonsum62  
-TRANS=human40.ccl -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MATEL=200000000 -USER=US10089514@cgn\_1.1.920@runat\_04102005.105745.8055  
-NCPU=6 -ICPU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEX=7

## Database :

Published Applications\_NA:\*  
1: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/prodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq:\*  
5: /cgn2\_6/prodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/prodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/prodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/prodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/prodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/prodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/prodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/prodata/1/pubpna/US10D\_PUBCOMB.seq:\*  
17: /cgn2\_6/prodata/1/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgn2\_6/prodata/1/pubpna/US10F\_PUBCOMB.seq:\*  
19: /cgn2\_6/prodata/1/pubpna/US10G\_PUBCOMB.seq:\*  
20: /cgn2\_6/prodata/1/pubpna/US10H\_PUBCOMB.seq:\*  
21: /cgn2\_6/prodata/1/pubpna/US10I\_PUBCOMB.seq:\*  
22: /cgn2\_6/prodata/1/pubpna/US10J\_NEW\_PUB.seq:\*  
23: /cgn2\_6/prodata/1/pubpna/US11A\_PUBCOMB.seq:\*  
24: /cgn2\_6/prodata/1/pubpna/US11\_NEW\_PUB.seq:\*  
25: /cgn2\_6/prodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
26: /cgn2\_6/prodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysts of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	1635	100.0	969	20	US-10-472-587-5 Sequence 5, Appl1
2	1588	97.1	3305	14	US-10-126-927-68 Sequence 68, Appl1
3	1588	97.1	12391	14	US-10-126-927-67 Sequence 67, Appl1
4	237	14.5	1122	16	US-10-137-310-3 Sequence 3, Appl1
5	233	14.3	1122	19	US-10-433-556-18 Sequence 18, Appl1
6	233	14.3	1238	16	US-10-380-133-7 Sequence 7, Appl1
7	225.5	13.8	1122	16	US-10-137-310-1 Sequence 1, Appl1
8	225.5	13.8	1122	18	US-10-634-548-16 Sequence 16, Appl1
9	225.5	13.8	1152	17	US-10-433-556-20 Sequence 20, Appl1
10	216	13.2	1830121	17	US-10-329-670-1 Sequence 1, Appl1
11	216	13.2	1830121	20	US-10-158-865-1 Sequence 1, Appl1
12	216	13.2	1830121	22	US-10-981-687-1 Sequence 1, Appl1
13	213	13.0	1792	22	US-10-450-763-1605 Sequence 1605, Ap
14	174.5	10.7	2731748	19	US-10-297-465A-1 Sequence 1, Appl1
15	165	10.1	2025	17	US-10-369-493-37621 Sequence 37621, A
16	159	9.7	88421	9	US-09-976-059-1 Sequence 1, Appl1
17	153.5	9.4	954	19	US-10-437-963-101181 Sequence 101181,
18	152	9.3	891	19	US-10-375-039-13 Sequence 13, Appl1
19	151	9.2	9025608	15	US-10-156-761-1 Sequence 1, Appl1
20	150.5	9.2	1491	15	US-10-156-761-1868 Sequence 1868, Ap
21	150.5	9.2	36602	21	US-10-762-107-1 Sequence 1, Appl1
22	150.5	9.2	9025608	15	US-10-156-761-1 Sequence 1, Appl1
23	150	9.2	42291	21	US-10-485-710-4 Sequence 2, Appl1
24	149	9.1	1236	21	US-10-500-416-2 Sequence 2, Appl1
25	149	9.1	1023	21	US-10-898-775-27 Sequence 27, Appl1
26	148	9.1	1023	22	US-10-911-160-21 Sequence 21, Appl1
27	147.5	9.0	1212	19	US-10-437-963-86432 Sequence 86432, A
28	147.5	9.0	1212	11	US-09-758-759-182 Sequence 182, App
29	147.5	9.0	37116	17	US-10-107-431-279 Sequence 279, App
30	147.5	9.0	2120	22	US-10-499-065A-239 Sequence 239, App
31	146	8.9	2280	22	US-10-499-065A-242 Sequence 242, App
32	146	8.9	2331	22	US-10-499-065A-244 Sequence 244, App
33	146	8.9	2457	22	US-10-499-065A-238 Sequence 238, App
34	146	8.9	2489	22	US-10-499-065A-237 Sequence 237, App
35	146	8.9	2520	22	US-10-499-065A-236 Sequence 236, App
36	146	8.9	2641	22	US-10-499-065A-243 Sequence 243, App
37	146	8.9	18435	15	US-10-156-761-412 Sequence 412, App
38	146	8.9	100000	15	US-10-156-761-15103 Sequence 15103, A
39	145	8.9	3957	16	US-10-200-562-193 Sequence 193, App
40	145	8.9	3957	16	US-10-237-451-193 Sequence 193, App
41	145	8.9	3957	22	US-10-945-050-193 Sequence 193, App
42	145	8.9	154746	10	US-09-827-688-8 Sequence 8, Appl1
43	145	8.9	154746	10	US-09-827-688-8 Sequence 8, Appl1
44	145	8.8	25617	21	US-10-485-710-7 Sequence 7, Appl1
45	143.5	8.8	25617	21	US-10-485-710-7 Sequence 7, Appl1

## ALIGNMENTS

RESULT 1  
US-10-472-587-5  
; Sequence 5, Application US/10472587  
; Publication No. US20040214274A1  
; GENERAL INFORMATION:  
; APPLICANT: YANAI, Naomi  
; APPLICANT: SUMIDA, Naomi  
; APPLICANT: WATANABE, Manabu  
; APPLICANT: MORIYA, Tatsuki  
; APPLICANT: MURAKAMI, Takehiko  
; TITLE OF INVENTION: Transformants Producing Substances P1022 Derivatives, Methods for  
; FILE REFERENCE: 2003-1302A/MMC/00144  
; CURRENT APPLICATION NUMBER: US/10/472,587  
; CURRENT FILING DATE: 2003-09-22  
; PRIOR APPLICATION NUMBER: 82227/2001



	PRIOR FILING DATE: 2001-03-22	
	NUMBER OF SEQ ID NOS: 40	
	SOFTWARE: PatentIn Ver. 2.1	
	SEQ ID NO 5	
	LENGTH: 969	
	TYPE: DNA	
	ORGANISM: Streptomyces venezuelae	
	FEATURE:	
	NAME/KEY: CDS	
	LOCATION: (1)..(966)	
	US-10-472-587-5	
	Alignment Scores:	
	Pred. No.: 1,65e-158	Length: 969
	Score: 1635..00	Matches: 322
	Percent Similarity: 100.00%	Conservative: 0
	Best Local Similarity: 100.00%	Mismatches: 0
	Query Match: 100.00%	Indels: 0
	DB: 20	Gaps: 0
US-10-089-514-6 (1-322) x US-10-472-587-5 (1-966)		
OY	1 MetSerGlyPheProArgSerValValAlaGlySerGlyValAlaGlyMetPhe	20
Db	1 ATGACGGCTTCCCGCAGCTGTCTGTGGCGGAGCGGGCGGTGGGCATGTTTC	60
OY	21 AlaGlyLeuLeuArgGlnAlaGlySerArgThrLeuValValAspLeuValProProPro	40
Db	61 GCCGGGCTGCTCGGGAGCGGGCAACGCCCATGCTGTGATCGACTCGAACGCCGCCG	120
OY	41 GlyArgProAspAlaCysLeuValGlyAspValThrAlaProGlyProGluLeuAlaAla	60
Db	121 GGACGGCGCGAGCGCTGCTGGGGCGAAGTCACCGCGCGCGGCCGGAACTCGCGGCC	180
OY	61 AlaLeuArgAspAlaAspLeuValLeuLeuAlaValHisGlnAspValAlaLeuLysAla	80
Db	181 GCGCTTCGGGAGCGGAGCTCGTCTGCTCGCGTAACAAGAGACGTGGCCCTCAAGGCC	240
OY	81 ValAlaProValThrArgLeuMetArgProGlyValAlaLeuLeuAlaAspThrLeuSerVal	100
Db	241 GTGGGCGCCGTACCCGGCTCATGGCGCGGGCGGCGTGTCGCCACACCCTGTCCGTC	300
OY	101 ArgThrGlyMetAlaAlaGluLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeu	120
Db	301 CGGACGGCGCATGGCGCGAGCTCGGGGCCAACGCCCGCGCTCCAGCACGTGGGCTTC	360
OY	121 AspPrometPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValThr	140
Db	361 AACCCGATGTTGCCCGCCCCCGCGGCAAGACCGGCCCTGTGGCGCGGTGCTACCC	420
OY	141 ArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGlnGlyGlyArgPro	160
Db	421 AGGACGGGCGCGGCGTCAACGCGCTCTGCGGCTGTGTGAGGGCGGCGGCGAGGCC	480
OY	161 ValArgLeuThrAlaGlnGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHis	180
Db	481 GTAACGGCTCAACGGGAGGAGACGACGACGAGCGGCGCACCCAGGCGCTGACGAC	540
OY	181 AlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeu	200
Db	541 GCGGTGCTCTCTCTCGGGCTCGGCTCGCGCGCTCGGCGCTGCAAGTCCGGGCTTG	600
OY	201 AlaAlaThrAlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGly	220
Db	601 GCGGGGACGGACCGCGCGCCCAACGAGTGTGCTGCGCTCTGTGGCGGTGTGCTGGC	660
OY	221 GlySerProGluValTyrgIyaSpirLegIArgSerAsnProArgAlaAlaSerAlaArg	240
Db	661 GGCACCGCCGAGGTGTACGGGGACATCGAGGTTCAAACCCCGGGCGCGTCCGGCGCC	720
OY	241 ArgAlaLeuAlaGluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArg	260
Db	721 CGGGGCTGTCCGAGGCGCTGTGCTCTTTCGCGCGCTGTGTGCGACGACCGGACCGT	780

QY	261	AlaAspAlaPProGlyValAGAlaAspAlaPProGlyHisProGlyGlyCyAspArgValAGly	280
Db	781	GCCACGCCCCCGGGGGCGCGGACGCCCGCGGCATCCGGGGATATCGACGGGCGGGG	840
QY	281	AsnLeuAspGlyValAPheGlyGlyLeuArgArgLeuMetGlyProGlyLeuAlaAlaGly	300
Db	841	AACTTCAGCGCGCTTCGGGGAACTCCGCGGCTATGGGACCGAGGCTTCGCGCGGGC	900
QY	301	GlnAspHisCyGlyGlyLeuPheArgThreLeuHisArgThraAspArgGlyGlyLys	320
Db	901	CAGACCACTGTCAGAGAGCTGTTCCGACCTCCACCGCACCGACGAGGAGGAG	960
QY	321	AspArg 322	
Db	961	GAACGA 966	
RESULT 2			
US-10-126-927-68			
; Sequence 68, Application US/10126927			
; Publication No. US20030082575A1			
; GENERAL INFORMATION:			
; APPLICANT: The Scripps Research Institute			
; APPLICANT: Schultz, Peter G			
; APPLICANT: Wang, Lei			
; APPLICANT: Anderson, John C			
; APPLICANT: Chin, Jason			
; APPLICANT: Liu, David R			
; APPLICANT: Magliery, Thomas			
; APPLICANT: Meggers, Eric L			
; APPLICANT: Mehl, Ryan A			
; APPLICANT: Pasternak, Miro			
; APPLICANT: Santoro, Stephen W			
; APPLICANT: Zhang, Zhiwen			
; TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids			
; FILE REFERENCE: 54-000120US			
; CURRENT APPLICATION NUMBER: US//10/126,927			
; CURRENT FILING DATE: 2002-04-19			
; PRIOR APPLICATION NUMBER: US 60/285,020			
; PRIOR FILING DATE: 2001-04-19			
; PRIOR APPLICATION NUMBER: US 60/355,514			
; PRIOR FILING DATE: 2002-02-06			
; NUMBER OF SEQ ID NOS: 79			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 68			
; LENGTH: 3305			
; TYPE: DNA			
; ORGANISM: Streptomyces venezuelae			
US-10-126-927-68			
Alignment Scores:			
Pred. No.: 4.64e-153 Length: 3305			
Score: 1588.00 Matches: 315			
Percent Similarity: 96.65% Conservative: 2			
Best Local Similarity: 96.04% Mismatches: 5			
Query Match: 97.13% Indels: 6			
DB: 14 Gaps: 1			
US-10-089-514-6 (1-322) x US-10-126-927-68 (1-3305)			
QY	1	MetSerGlyPheProArgSerValValAlaGlyGlySerGlyAlaValAGlyGlyMetPhe	20
Db	2007	ATGAGCGGCTTCCCGCGAGCGTGTCTGCGCGGACCGGAGCGGTGGCGGCATGTTTC	2066
QY	21	AlaGlyLeuLeuArgGlyAlaGlySerArgThrLeuValValAspLeuValProProPro	40
Db	2067	GCCGGGCTGTGTGCGGAGCGGCGGACGCGACGCTGTCTGCACTGTAACGCGCGCG	2126
QY	41	GlyArgProAspAlaCyLeuLeuValGlyAspValThrAlaProGlyProGlyLeuAlaAla	60
Db	2127	GGAGCGCGGACGCGCTTCTGTGTGGGCGACGTACCGCGCGGGGCCCGAGCTTCGCGGC	2186
QY	61	AlaLeuArgAspAlaAspLeuValLeuLeuAlaValHisGlyLeuAspValAlaLeuLysAla	80







```
QY 241 ARGAlaleuAlaIleuAArgSerPheAlaIleuValaIyAsp----- 256
Db 3081 CGGGCCCTCGCCAGAGCCCTGCGCTCTTCCGCGCGCTGATCGGCACGACCCGACCGC 3140
QY 257 -----AspProspaRrgAlaAspAlaProGlyAArgAlaAspAlaProGlyYHISProGly 274
Db 3141 GCCGAGAACCCGAGCGCGCGCGACGACCCGACCGACCGACCAACCCGCGCATCCCGGG 3200
QY 275 GlyCyAspGlyValaIleuAAspGlyValaPheGlyGlyleuAArgArgleuMetGly 294
Db 3201 GGAATGGAGCGCGCCGCGGAACCTCGACGCGCTTTCGAGAGAACTCCGCGCGCTCATGGCA 3260
QY 295 ProGlyleuAlaIleuAArgGlyAspHisCySGlyGlyleuAArgThreuhISArgThr 314
Db 3261 CCGGAGCTCGCGCGCGCGACCACTGCGCAGAGCTGTTCGCACTCCGACCCGACCGCACCC 3320
QY 315 AspaSPGlyGlyIlyAspArg 322
Db 3321 GACGACGAAGGCGAGAGACCGA 3344

RESULT 4
US-10-137-310-3
; Sequence 3. Application US/10137310
; Publication No. US20030176675A1
; GENERAL INFORMATION:
; APPLICANT: Valentim, Henry E.
; APPLICANT: Mitsky, Timothy A.
; TITLE OF INVENTION: Tyra Genes and Uses Thereof
; FILE REFERENCE: 16515.147
; CURRENT APPLICATION NUMBER: US/10/137, 310
; CURRENT FILING DATE: 2002-05-03
; PRIOR APPLICATION NUMBER: US 60/289,527
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1122
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-137-310-3

Alignment Scores:
Pred. No.: 1,32e-14 Length: 1122
Score: 237.00 Matches: 78
Percent Similarity: 45.39% Conservative: 50
Best Local Similarity: 27.66% Mismatches: 110
Query Match: 14.50% Indels: 44
DB: 16 Gaps: 8

US-10-089-514-6 (1-322) x US-10-137-310-3 (1-1122)
QY 8 ValValaIaIyGlySerGlyAlaValaIyGlyMetPheAlaIleuAArgGlyAla 27
Db 304 GTTATGCTTGGCGCGCGCGCTCAGATGGAGCGTCTGTTCGAGAGATGCTGACACTCTCG 363
QY 28 GlySerArgThreuhValaIleuAAspValaProProGlyAArgProAspAlaCysleu 47
Db 364 GGTATCAGGTGCGGATTCGAG-----CAACATGACTGGATTCGA 405
QY 48 ValGlyAspValThraIaProGlyProGlyleuAlaIleuAArgAspAlaAspleu 67
Db 406 GCGGCTGATGTTGTTGCC-----GATGCCGAATG 435
QY 68 ValIleuAlaIaIleuAHisGlyAspValaIleuAAspValaIleuAProValThraArgleu 87
Db 436 GTGATGTAGTGTGCCAATCCAGTTACTGAGCAAGTTATGGCAAAATTACCGCCTTGA 495
QY 88 MetArgProGlyAlaIleuAlaAspThreuhSerValaArgThrgly--MetAlaAla 106
Db 496 CGGAAG--GATTGTATCTCGTGAATCGGCATCTGTAATAATGACCATTAACGGCC 552
QY 107 GluIleuAlaIaIleuAProGlyValGlnHisValGlyLeuAsnProMetPheAlaPro 126
```

```
Db 553 ATGTCGCGCGCGACAGATGCGCGGTA-----CTGGGGTTACACCCAAATGTTGGTCCG 606
QY 127 AlaAlaGlyMetThrcIyArgProValaIaIaValaIaThraArgAspGlyProGlyVal 146
Db 607 GACAGCGGTAGCTTGGCAAAAGCAAGT-----GTGGTCTGGGTGATGAGACTAAACCG 660
QY 147 ThrAla-----LeuIleuAArgleuValGlyGlyGlyValaArgProValaArgleu 163
Db 661 GAAGCATACCAATGTTTCTGGAGCAAAATTCAAGTCTGGGGCGCTGTCATCGTATT 720
QY 164 ThrAlaGlyGlnHisAspArgThraIaIaIaThrcIleuAAlaIleuThraIaIleu 183
Db 721 AGGCGCTGACACACATCAATATAGCGTTATTCAGGCACTGGCCACTTGGCTACT 780
QY 184 LeuSerPheGlyLeuAlaIleuAlaArgleuGlyValaAspValaArgAlaIleuAlaIthr 203
Db 781 TTTCCTTACGGGCTGCACCTGGCAGAAAGAAATGTTCACTTGAGCAATTTCTGGCGCTC 840
QY 204 AlaProProHisGlnValIleuAlaIleuAlaIleuAlaArgValIleuGlySerPro 223
Db 841 TCTTCGCGATTATTAACCGCTTGAGCTGGCGATGTCGGGCGACTGTTCCTCAGATCCG 900
QY 224 GluValaIyGlyAspIleuAArgSerAsnProAArgAlaIaIaSerAla----- 239
Db 901 CAGCTTATCCGACATTAATTATGTCTCAGAGCGTTAATCTGACCAACCGTTAC 960
QY 240 -----ArgArgAlaIleuAla 244
Db 961 TATTAAGCTTGGCGGACGAGTTAGTTGCTGAGCAGGCGGATTAAGCGCTTATT 1020
QY 245 GluAlaIleuAArgSerPheAlaIleuValaIyAspAspProAspArgAlaAspAlaPro 264
Db 1021 GACAGTTCCGGAAGGTGAGACACTGTCGGGATTTACGACAGGCTTTTCAGAGTGA 1080
QY 265 GlyArg 266
Db 1081 AGCCGC 1086

RESULT 5
US-10-433-556-18
; Sequence 18. Application US/10433556
; Publication No. US20040117872A1
; GENERAL INFORMATION:
; APPLICANT: Matringe, Michel
; APPLICANT: Rippert, Pascal
; TITLE OF INVENTION: Novel Targets For Herbicides And Transgenic Plants Resistant To
; FILE REFERENCE: 5500*120
; CURRENT APPLICATION NUMBER: US/10/433,556
; CURRENT FILING DATE: 2003-06-04
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
; LENGTH: 1122
; TYPE: DNA
; ORGANISM: Escherichia coli
; NAME/KEY: CDS
; LOCATION: (1)..(1122)
US-10-433-556-18

Alignment Scores:
Pred. No.: 3.4e-14 Length: 1122
Score: 233.00 Matches: 77
Percent Similarity: 45.04% Conservative: 50
Best Local Similarity: 27.30% Mismatches: 111
Query Match: 14.25% Indels: 44
DB: 19 Gaps: 8

US-10-089-514-6 (1-322) x US-10-433-556-18 (1-1122)
QY 8 ValValaIaIyGlySerGlyAlaValaIyGlyMetPheAlaIleuAArgGlyAla 27
```



```
Db 304 GTTATCGTCGGCGGTGCGGTACAGTGGAGCGCTGTTCCAGAAAGATGCTGACCCCTTCG 363
Qy 28 GlysSerArgThrLeuValValAspLeuValProProProGlyArgProAspAlaCysLeu 47
Db 364 GGTATTCAGGTGGCGATTCTGGAG-----CAACATGACTGGGATCGA 405
Qy 48 ValGlyAspValThrAlaProGlyProGluLeuAlaAlaLeuArgAspAlaAspLeu 67
Db 406 GCGGCTGATTTGTGTC-----GATGCCGGGAATG 435
Qy 68 ValLeuLeuAlaValHisGluAspValAlaLeuValAlaValAlaProValThrArgLeu 87
Db 436 GTGATTGTAGTGTGGCAATCCAGTTACTGAGCAAGTATTGGCAATTAACGCCCTTTA 495
Qy 88 MetArgProGlyAlaLeuAlaAspThrLeuSerValArgThrGly---MetAlaAla 106
Db 496 CCGAAA---GATTGTATTCGTGTCGATCTGGCAATCAATGAAATGGCCATTAACAGGCC 552
Qy 107 GluLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeuAsnProMetPheAlaPro 126
Db 553 ATGCTGCTGGCGCATGATGTCGGTG-----CTGGGCTTACCCGATGTTCCGTCCG 606
Qy 127 AlaAlaGlyMetThrGlyArgProValAlaAlaValAlaThrArgAspGlyProGlyVal 146
Db 607 GACAGCGGTAGCTGCGCAAGCAAGTT-----GTGCTGTGTCGTGATGACGTAACCG 660
Qy 147 ThrAla-----LeuLeuArgLeuValGluGlyGlyGlyValArgProValArgLeu 163
Db 661 GAAGCATACCAATGTTCTTGAGAGCAATTCAGGTCGGGGCGCTCGGTCGATCGTATT 720
Qy 164 ThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValAla 183
Db 721 AGCGCCGTCAAGCAGCATCGAATATGCGCTTATTACAGCACTGCCCATTTGCTACT 780
Qy 184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaAlaThr 203
Db 781 TTTCGTTACGGGCTGACCTGGCGAAGAAATGTTCACTGACCTGCAACTTTCGGGCTC 840
Qy 204 AlaProProProHisGlnValLeuLeuAlaLeuAlaLeuValLeuGlySerPro 223
Db 841 TCTTCGCGATTACCCCTTGAGCTGCGATGTCGGGCGACCTGTTGCTCAGATCCG 900
Qy 224 GluValTyrGlyAspIleGlnArgSerAsnProArgAlaAlaSerAla----- 239
Db 901 CAGCTTATGCCGACATCATTAATGTCGTACAGCGTAATCTGGGTTAATCAACGTTAC 960
Qy 240 -----ArgArgAlaLeuAla 244
Db 961 TATTAAGGTTTCGGCGAGCGATTGACTGCTGAGCAGGGCCATAGCAGGGCTTATT 1020
Qy 245 GluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArgAlaAspAlaPro 264
Db 1021 GACAGTTTCGCAAGGTGAGACACTGGTTCGGCGATTACGACAGCGTTTTCAGAGTAA 1080
Qy 265 GlyArg 266
Db 1081 AGCCGC 1086
```

```
RESULT 6
US-10-380-132-7
; Sequence 7, Application US/10380132
; Publication No. US20030182679A1
; GENERAL INFORMATION:
; APPLICANT: Sungenix GmbH & Co. KGaA
; TITLE OF INVENTION: Improved processes for vitamin E biosynthesis
; FILE REFERENCE: NAE445/2000
; CURRENT APPLICATION NUMBER: US/10/380,132
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1238
; TYPE: DNA
```

```
; ORGANISM: Escherichia coli
;
; FEATURE:
; NAME/KEY: gene
; LOCATION: (7)..(1232)
; OTHER INFORMATION: CysA gene coding for bifunctional chorismate
; OTHER INFORMATION: mutase / prephenate dehydrogenase
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (25)..(1143)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(6)
; OTHER INFORMATION: restriction site linker
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1233)..(1238)
; OTHER INFORMATION: restriction site linker
US-10-380-132-7
```

## Alignment Scores:

Pred. No.:	3,81e-14	Length:	1238
Score:	233.00	Matches:	77
Percent Similarity:	45.04%	Conservative:	50
Best Local Similarity:	27.30%	Mismatches:	111
Query Match:	14.25%	Indels:	44
DB:	16	Gaps:	8

US-10-089-514-6 (1-322) x US-10-380-132-7 (1-1238)

```
Qy 8 ValValValGlyGlySerGlyAlaValAlaGlyMetPheAlaGlyLeuLeuArgGluAla 27
Db 328 GTTATCGTCGGCGGTGCGGTGCGATGATGAGACCGCTGTCAGAAAGATGCTGACCCCTTCG 387
Qy 28 GlysSerArgThrLeuValValAspLeuValProProProGlyArgProAspAlaCysLeu 47
Db 388 GGTATTCAGGTGCGGATTCGAG-----CAACATGACTGGGATCGA 429
Qy 48 ValGlyAspValThrAlaProGlyProGluLeuAlaAlaLeuArgAspAlaAspLeu 67
Db 430 GCGGCTGATTAATTTGTC-----GATGCCGGGAATG 459
Qy 430 GCGGCTGATTAATTTGTC-----GATGCCGGGAATG 459
Db 68 ValLeuLeuAlaValHisGluAspValAlaLeuValAlaValAlaProValThrArgLeu 87
Qy 460 GTGATTGTAGTGTGCCAATCCAGTTACTGAGCAAGTATTGGCAATTAACGCCCTTTA 519
Db 88 MetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrGly---MetAlaAla 106
Qy 520 CCGAAA---GATTGTATTCGTGTCGATCTGGCATCGAATAAATGGCCATTAACAGGCC 576
Db 107 GluLeuAlaAlaHisAlaProGlyValGlnHisValAlaGlyLeuAsnProMetPheAlaPro 126
Qy 577 ATGCTGTCGGCGATATGTCGGTG-----CTGGGCTTACCCGATGTTCCGTCCG 630
Db 127 AlaAlaGlyMetThrGlyArgProValAlaAlaValAlaThrArgAspGlyProGlyVal 146
Qy 631 GACAGCGGTAGCTGCGCAAGCAAGTT-----GTGCTGTGTCGTATGACGTAACCG 664
Db 147 ThrAla-----LeuLeuArgLeuValGluGlyGlyGlyValArgProValArgLeu 163
Qy 685 GAAGCATACCAATGTTCTTGAGAGCAATTCAGGTCGGGGCGCTCGGTCGATCGTATT 744
Db 164 ThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValAla 183
Qy 745 AGCGCGGTGAGCAGCATCAATGATGCGTTTATTACAGCACTGCCCATTTGCTACT 804
Db 184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaAlaThr 203
Qy 805 TTTCGTTACGGGCTGACCTGCGCAAGAAATGTTCACTTACGCAACTTTCGGGCTC 864
Db 204 AlaProProProHisGlnValLeuLeuAlaLeuAlaArgValLeuGlySerPro 223
Qy 865 TCTTCGCCGATTTAACCGCTTGAGCTGCGGATGTCGGGCGAATGTTGCTCAGATCCG 924
```



QY	224	GIUValaIYrGlaYAspLIgLIaRgSerIaSPrOaRglaIaLaSerAa-----	233
Db	925	CAGCTTTATGCCGACATCATTTATGTCGTGAGCGCTAACTGGCGTTATCAACAGCTTAC	984
QY	240	-----ArgRglaLeuLa	244
Db	985	TATAAGCGTTTGGCGAGGCGCATGTAGTTCTCGAGACAGGCGGATTAAGAGCGCTTTATT	1044
QY	245	GIUaLaLeuYrSerPhealAaLaLeuValGIYAspAspProAspRglaIaAspLaPro	264
Db	1045	GACAGTTTCCCGCAAGATGGAGACACTGTTCGCGCATTAACGACACAGCGTTTTCAGATGA	1104
QY	265	GIYArG	266
Db	1105	AGCCGC	1110
RESULT 7			
US-10-137-310-1			
; Sequence 1, Application US/10137310			
; Publication No. US20030176675A1			
GENERAL INFORMATION:			
; APPLICANT: Valentin, Henry E.			
; APPLICANT: Miteky, Timothy A.			
; TITLE OF INVENTION: Tyra Genes and Uses Thereof			
; FILE REFERENCE: 16515.147			
; CURRENT APPLICATION NUMBER: US/10/137,310			
; PRIOR FILING DATE: 2002-05-03			
; PRIOR APPLICATION NUMBER: US 60/289,527			
; PRIOR FILING DATE: 2001-05-09			
; NUMBER OF SEQ ID NOS: 12			
; SOFTWARE: PatentIn version 3.0			
; SEQ ID NO 1			
; LENGTH: 1122			
; TYPE: DNA			
; ORGANISM: Erwinia herbicola			
US-10-137-310-1			
Alignment Scores:			
Pred. No.: 2,01e-13			
Score: 225.50			
Percent Similarity: 48.26%			
Best Local Similarity: 30.43%			
Query Match: 13.79%			
DB: 16 Gaps: 6			
US-10-089-514-6 (1-322) x US-10-137-310-1 (1-1122)			
QY	8	ValValValGIgYISerGIYAlaValGIgIYmetPhealGIYLeuLeuArGIUaLa	27
Db	304	GTGATTGTCGGTGTGAAGGCGCAATGGCGCGCTGTTGAAAAAATGCTCGGCGTATCA	363
QY	28	GIYSerArGThrLeuValValAspLeuValProProProGIYArGProAspLaIcYsLeu	47
Db	364	GAGCTACACGGTTAAAAAGCTGGAT-----	387
QY	48	ValGIAspValThrAlaProGIYProGIULeuLaIaLaLeuArGAspAlaAspLeu	67
Db	388	-----AAAGAGAGACTGGCTCAGGCTGAGACTCGCTCAGGATGCGCGGATG	435
QY	68	ValLeuLeuLaValHisGIuAspValAlaLeuYsAlaValAlaProValThrArgLeu	87
Db	436	GTGATCTTACCGTGGCCGATTCACCTGACCGAGCGATGATGGCCCAATGCCACCACTG	495
QY	88	MetArGProGIYAlaLeuLeuLaAspThrLeuSerValArgThr---GIYMetAlaLa	106
Db	496	CCGGAA---GATTGTATTCTGGTCATCTGGCGCTAGTCAAAAACCGGCTCTGAGGCA	552
QY	107	GIULeuLaIaLaHisAlaProGIYValGIuHisValGIYLeuAspProMetPheAlaPro	126
Db	553	ATGCTGGCTGCCCATTAACGGGCGCTGA-----CTGGGCTCGCATCCGATGTTTGGCCG	606
QY	127	AlaAlaGIYmetThrGIYArGProValAlaLaValValThrArGAspGIYProGIYVal	146

Df		607	GACAGCGGACGGCTGGCAAAACAAGTG-----GTGTCCTGTGTATGTAAGAACAACC	668
Oy		147	ThralA-----LeueuarlgleualgluglygiylgYargProvalArgleu	163
Dd		661	GAAGCCGATCAGTGGTTCTCGAGGCAGATTACAGTCTGGGGTGGCGGTCTGCATCGTATC	729
Oy		164	ThrAlagIunhiSaSparghrThrralaaLmrglnalaleturHlAlavalleu	183
Dd		721	AGCGCTGTGAGCATCACCAAACAATGCCATTTCATTCAGGGCGCTGCCTACTTGTCCACC	786
Oy		184	LenseRhegliyeulaaleuualaargleugiylaapValArgalAleuaLaLAthr	203
Dd		781	TTCGCTTAGTGTTCGATTTCATTCAGCCAAGAACGTCATTCGATCACTGCCCCCTC	840
Oy		204	AlaePRePhroHiSiInValleuleuualaleuualaargValleugiylSerPro	223
Dd		841	TGTCGCCCATTACCGCGCTTGAACTGGGAGATGNGGGCGGTGTTCGCTCAGATCCG	900
Oy		224	GlualVtyrcLYasPllegImSrSen	233
Dd		901	CAACTCATGCCGATTCATCATGTCTTCA	930
 RESULT 8 US-10-634-548-16				
/	Sequence 16,	Application US/10634548		
/	Publication NO.	US20040045051AI		
/	GENERAL INFORMATION:			
/	APPLICANT:	No. US20040045051AIRis, Susan R		
/	APPLICANT:	Lincoln, Kim		
/	APPLICANT:	Abad, Mark Scott		
/	APPLICANT:	Eilers, Robert		
/	APPLICANT:	Hartweyer, Karen Kindle		
/	APPLICANT:	Hirsberg, Joseph		
/	APPLICANT:	Karnanandas, Balasuojini		
/	APPLICANT:	Moshiri, Farhad		
/	APPLICANT:	Stein, Joshua C.		
/	APPLICANT:	Valetnin, Henry E.		
/	APPLICANT:	Venkatesh, Tyamagondlu V.		
/	TITLE OF INVENTION:	To copropherol biosynthesis related genes and used thereof		
/	FILE REFERENCE:	Ren-01-125		
/	CURRENT APPLICATION NUMBER:	US/10/634,548		
/	CURRENT FILING DATE:	2003-08-05		
/	PRIOR APPLICATION NUMBER:	us 60/400,689		
/	PRIOR FILING DATE:	2002-08-05		
/	NUMBER OF SEQ ID NOS:	79		
/	SOFTWARE:	Patentin version 3.2		
/	SRO ID NO	16		
/	LENGTH:	1122		
/	TYPE:	DNA		
/	ORGANISM:	Erwinia herbicola		
/	US-10-634-548-16			
 Alignment Scores: Pred. No.:            2,01e-13          Length:         1122				
Score:              225.50           Matches:         70				
Percent Similarity: 48..26%          Conservative:    41				
Best Local Similarity: 30..43%          Mismatches:    94				
Query Match:        13..79%          Indels:          25				
DB:                   18           Gaps:            6				
 US-10-089-514-6 (1-322) x US-10-634-548-16 (1-1122)				
Oy		8	VaiVaIVaIGlVGISerGIYaIVaIGlyMeCPheaIgLyLeuLeUAgrGLUaIA	27
Dd		304	GTATTGTGCGTGGTAAGGCGCAKMTGGGCGCGCTTTGAAAAAATGCTCGGGCTRCA	363
Oy		28	GIYSerARgtThrlauValaspHeuvAlPrObRoFrogIyaGrPRoApalAcysleu	47
Dd		364	GGCTACACGGTTAAAACGCTGAT-----	387
Oy		48	VaIGLYAsPVaIThRALPrOGILyPROGLIUeULaalALeuARgsPaIALaPLEu	67
Dd		388	-----AAAGAGCACTGGCTCAGAGCTGAGCTTGCTCAGCGATGGCGGATG	435



QY 68 ValValLeuLeuAlaValHisGluAspValAlaLeuLysAlaValAlaProValThrArgLeu 87  
 |||.....  
 Db 436 GTGATGATTAAAGCCGCGATTCACTGACCGAGCAGGTGATGCCAATGCCACCACTG 495  
 QY 88 MetArgProGlyValLeuLeuAlaAspThrLeuSerValArgThr---GlyMetAlaAla 106  
 :|||.....  
 Db 496 CCGGAA---GATTGTATTCTGTCGTGATCTGCGCGTCAATAAAACCGGCCCTGTGAGGCA 552  
 QY 107 GLULeuAlaAlaHisAlaProGlyValGlnHisValGlyLeuAsnProMetPheAlaPro 126  
 |||.....  
 Db 553 ATGCTGCGCTGCCCATTAACGGGCGCTGTA-----CTGGGTCGTGATCGATGTTGGCCG 606  
 QY 127 AlaAlaGlyMetThrGlyArgProValAlaAlaValAlaThrArgAsnGlyProGlyVal 146  
 607 GACAGCGGACGCTGGCAAAACAGTG-----GTGGTCTGCTGTGATGAAAGCAACACCG 660  
 Db 147 ThrAla-----LeuLeuArgLeuValGluGlyGlyGlyArgProValArgLeu 163  
 |||.....  
 QY 661 GAAGCGTATCAGTGGTTCCTGGAGCAGATTCAGTCTGGGGTGGCGCTGTCATCGTATC 720  
 Db 164 ThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeu 183  
 :|||.....  
 QY 721 AGCGCTGTGAGCATGACCAACATGCGATTCTTCAGGCGCGCTGCTTGCTACC 780  
 Db 184 LeuSerPheGlyLeuValLeuAlaValArgLeuGlyValAspValArgAlaLeuAlaThr 203  
 781 TTCGCTTAATGTTGTCATTTAAGCCGAAGACATCTCATCTGATCACTGCGCTC 840  
 QY 204 AlaProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGlyGlySerPro 223  
 :|||.....  
 Db 841 TCGTCGCCATTATTCGGCGCTGAACGTGGCATGTGGGGGCTGTTCGCTCAGATCCG 900  
 QY 224 GluValTyrGlyAspIleGlnArgSerAsn 233  
 :|||.....  
 Db 901 CAACCTATGCGGATATCATCATGATCTTCA 930  
 RESULT 9  
 US-10-433-556-20  
 ; Sequence 20, Application US/10433556  
 ; Publication No. US20040117872A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Matrinage, Michel  
 ; APPLICANT: Ripperet, Pascal  
 ; TITLE OF INVENTION: Novel Targets For Herbicides And Transgenic Plants Resistant To  
 ; FILE REFERENCE: 5500\*120  
 ; CURRENT APPLICATION NUMBER: US/10/433,556  
 ; CURRENT FILING DATE: 2003-06-04  
 ; NUMBER OF SEQ ID NOS: 30  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 20  
 ; LENGTH: 1152  
 ; TYPE: DNA  
 ; ORGANISM: Erwinia herbicola  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(1119)  
 US-10-433-556-20  
 Alignment Scores:  
 Pred. No.: 2, 07e-13 Length: 1152  
 Score: 225.50 Matches: 85  
 Percent Similarity: 42.72% Conservative: 44  
 Best Local Similarity: 28.15% Mismatches: 119  
 Query Match: 13.79% Indels: 55  
 DB: 19 Gaps: 9  
 US-10-089-514-6 (1-322) x US-10-433-556-20 (1-1152)  
 QY 8 ValValValGlyGlySerGlyAlaValGlyGlyMetPheAlaGlyLeuLeuArgGluAla 27  
 |||.....  
 Db 304 GTGATTTGCTGTGTGAAGGCCAGATGGCGCGCTGTTGAAAAAAAGCTCGGGCTATCA 363

Oy 28 GYSerIarGhThrLeuValValAspLeuValProProProGlyAArgProAspAlaCylLeu 47  
 |||  
 Db 364 GGCTACACCGCTTAAACCGCTGGAT----- 387  
 Oy 48 ValGlyAspValThralaProGlyProGlyLeuValAlaAlaLeuAArgAspAlaAspLeu 67  
 |||  
 Db 388 -----AAGAGGACTGGCTTCAGGCTGAGACTCGCTCAGCGATCGCGGAATG 435  
 Oy 68 ValLeuLeuAlaValHisGlyAspValAlaLeuValAspAlaProValThraGlyLeu 87  
 |||  
 Db 436 GTGATCTATTAGCGCGCATTCACCTGACCGAGACAGGTGATTGGCCAACTGCCACCACTG 495  
 Oy 88 MetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThr---GlyMetAlaAla 106  
 |||  
 Db 496 CCGGAA--GATTGTATCTGGTCGATCTGGCGCTCAATCAAAACCGAGCTTCAGGCA 552  
 Oy 107 GluLeuValAlaHisAlaAspProGlyValGlnHisValGlyLeuAspProMetPheAlaPro 126  
 |||  
 Db 553 ATGCTGGCTGCCCTTAACGGGCTGTA-----CTGGGTCTGCATTCGAGTGTGGGCCG 606  
 Oy 127 AlaAlaGlyMetThrGlyArgProValAlaAlaValValThraArgAspGlyProGlyVal 146  
 |||  
 Db 607 GACAGCGGACCGCTGCGAAACAGGTG-----GTGCTGTGTGTGATGGAAGCAACCG 660  
 Oy 147 ThrAla-----LeuLeuArgLeuValGluGlyValGlyGlyArgProValArgLeu 163  
 |||  
 Db 661 GAACCGATATCATGCTGCTCGAGACAGATTAGCTGCGAGTGGCGCTTCGCATGTATC 720  
 Oy 164 ThrAlaGluGlnHisAspArgThrThralaAlaThrGlnAlaLeuThrHisAlaValLeu 183  
 |||  
 Db 721 AGCGCTGTGAGCATGACAGAACATGCGATTCATTACGGCGCTGCGCTTGTGCATC 780  
 Oy 184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaAlaThr 203  
 |||  
 Db 781 TTCCTTATGCTGTGCTATTAGCCGAAGAACGTCAATGTGATCAGCTGCGCTGCGCTC 840  
 Oy 204 AlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGlyGlySerPro 223  
 |||  
 Db 841 TCGTGCCCATTTTACCGGCTTTGAACCTGGCATGTGGGGCGGTGTTCCTCAGCAATCG 900  
 Oy 224 GluValTyrglyAspIleGlnArgSerAspProArgAlaAlaSerAlaArgAlaLeu 243  
 |||  
 Db 901 CAATCTATCGGATATCATCATCTCTTCAGAGGTA--TCGGCGCTGATTAACCTTA 959  
 Oy 244 AlaGluAlaLeu-----ArgSerPheAlaAlaLeuValGlyAspAspPro----- 258  
 |||  
 Db 960 TTACACACGGCTTTGGTGAACGATTCGCGCTGTGAACAGGCGACACAGCAGCGCTTAT 1019  
 Oy 259 -----AspArgAlaAspAlaProGlyArg 266  
 |||  
 Db 1020 CGCAGCTTTTAAACCGGTTGAACAGTGTGTGGCGATCAGCAAAACGTTCTGGTGA 1079  
 Oy 267 AlaAspAlaPro-----GlyHisProGly 274  
 |||  
 Db 1080 AAGCGGAAGCTGTGGATCGCGCATGACAGCGCCCATTAATAAAGCATCCAGTT 1139  
 Oy 275 GlyCys 276  
 |||  
 Db 1140 GGATGC 1145  
 RESULT 10  
 US-10-329-670-1  
 ; Sequence 1, Application US/10329670  
 ; Publication No. US20040018503A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fleischmann et al.  
 ; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag  
 ; FILE REFERENCE: PH186P1  
 ; CURRENT APPLICATION NUMBER: US/10/329,670  
 ; CURRENT FILING DATE: 2002-12-24  
 ; PRIOR APPLICATION NUMBER: US 09/643,990



```

? PRIOR FILING DATE: 2000-08-23
? PRIOR APPLICATION NUMBER: US 08/487,429
? PRIOR FILING DATE: 1995-06-07
? PRIOR APPLICATION NUMBER: US 08/426,787
? PRIOR FILING DATE: 1995-04-21
? NUMBER OF SEQ ID NOS: 1
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 1
? LENGTH: 1830121
? TYPE: DNA
? ORGANISM: Haemophilus influenzae
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (4747)..(4747)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (9921)..(9921)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (10150)..(10150)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (36543)..(36543)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (29298)..(29298)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (36551)..(36551)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (36636)..(36636)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (40808)..(40810)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (44416)..(44416)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (44905)..(44905)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (44975)..(44975)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (45593)..(45593)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (45732)..(45732)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (47036)..(47036)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (51334)..(51334)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
```

```

? LOCATION: (51602)..(51602)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (51786)..(51786)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (51805)..(51805)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (55369)..(55369)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (65309)..(65309)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (65313)..(65313)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (80024)..(80024)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (100091)..(100091)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (102696)..(102696)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (105121)..(105121)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (107248)..(107248)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (117136)..(117136)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (119750)..(119750)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (119924)..(119924)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (120038)..(120038)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (121344)..(121344)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (122167)..(122167)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (122336)..(122336)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (131340)..(131340)
```



```

OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (139510)..(139510)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152530)..(152530)
OTHER INFORMATION: n equals a, t, g or c

Alignment Scores:
Pred. No.: 9.5e-09 Length: 1830121
Score: 216.00 Matches: 77
Percent Similarity: 45.26% Conservative: 47
Best Local Similarity: 28.10% Mismatches: 118
Query Match: 13.21% Indels: 32
DB: 17 Gaps: 8

US-10-089-514-6 (1-322) x US-10-329-670-1 (1-1830121)
QY 2 SerGlyPheProArgSerValValValGlySerGlyValValGlyMetPheAla 21
DB 1370739 TCTGATATTACCAAAATTTGATTATGGCGGTTATGGCAAAATTAGCGGCTATTATGGC 1370798
QY 22 GlyLeuLeuArgGluIaGlySerArgThreValValValPheLeuValProProGly 41
DB 1370799 CGTTATTACGTGCTGCTATCCATTCTATTATTAGT----- 1370840
QY 42 ArgProAspAlaCysLeuValGlyAspValThrAlaProGlyProGluLeuAlaAlaAla 61
DB 1370841 CGCGAAGATTGGCGGCTGGAAGTATT----- 1370870
QY 62 LeuArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuValAlaVal 81
DB 1370871 TTAGCGAATGCTGATCGGATCGTTTCGTCCTATTATCTCACCTTAGAACAATT 1370930
QY 82 AlaProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerVal--- 100

```

```

DB 1370931 GAACGCTTAAACCTTATTAAACGAAACATGCTACTTGCAGATTAACTCTGTTAAG 1370990
QY 101 ArgThrGlyMetAlaIaGluLeuAlaAlaHisAlaAlaProGlyValGlnHisValGlyLeu 120
DB 1370991 CGTGAAACCGCTGCAAAATGCTGAAGTTCAATACGTGCTGCT---TTAGGTTTA 1371044
QY 121 AspProMetPhe---AlaProAlaIaGlyMetThrGlyArgProValAlaAlaValAla 139
DB 1371045 CATCCAAATGTTTGTCAGATATTCGACAGATATGCAAAACAAAGTGTTGCGCTTGAT 1371104
QY 140 ThrArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGluGlyGlyArg 159
DB 1371105 GGAAGTTTCCGTAACGTTATGAATGTTTACTTGAGCAAAATTAATTTGGGGTCAAAA 1371164
QY 160 ProValArgLeuThrAlaGluGluHisAspArgThrThrAlaAlaThrGlnAlaLeuThr 179
DB 1371165 ATTTATCAACCAATGCGCAGAACGATCATATATATGACTTATATACAGCTTGCGC 1371224
QY 180 HisAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAla 199
DB 1371225 CATTTTCCAGCTTTGCGAATGTTTACACTTTCACAAACGCCATTATCTCGCTAAT 1371284
QY 200 LeuAlaAlaThrAlaProProProHISGlnValLeuLeuAlaLeuLeuAlaArgValLeu 219
DB 1371285 TTATTGGCACTTCTTCCCTATTATTCGTTAGTACCTTGATAGTGGCTTATTT 1371344
QY 220 GlyGlySerProGluValIlyrGlyAspIle-----GlnArgSerAspProAlaAla 237
DB 1371345 GCGCAAGATGCAAGCTTTACGAGATATTATATGATATGATACAGAAATTTAGCGGTA 1371404
QY 238 -----SerAlaArgArgAlaLeuAlaGluAlaLeuArgSerPheAlaAlaLeuValGly 255
DB 1371405 ATTGAAACCGCTAAACCAACTTATCGACGAGGCTTAACTTCTTT----- 1371449
QY 256 AspAspProAspArgAlaAspAlaProGlyArgAlaAspAla 269
DB 1371450 -----GAAATATATGATGCTCAAGTTTATTGATGCT 1371482

RESULT 11
US-10-158-865-1
Sequence 1, Application US/10158865
Publication No. US20040203093A1
GENERAL INFORMATION:
APPLICANT: Fleischmann et al.
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Fra
FILE REFERENCE: PB186P2C1D1
CURRENT APPLICATION NUMBER: US/10/158,865
PRIOR FILING DATE: 2002-06-03
PRIOR APPLICATION NUMBER: US 09/557,884
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: US 08/476,102
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: US 08/426,787
PRIOR FILING DATE: 1995-04-21
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1830121
TYPE: DNA
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: misc feature
LOCATION: (4747)..(4747)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a,t,c, or g

```



```
?
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (29298)..(29298)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (36543)..(36543)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (36551)..(36551)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (36636)..(36636)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (40810)..(40810)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (44416)..(44416)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (44905)..(44905)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (44975)..(44975)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (45593)..(45593)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (45732)..(45732)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (47036)..(47036)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (51334)..(51334)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (51602)..(51602)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (51786)..(51786)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (51805)..(51805)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (55369)..(55369)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (65309)..(65309)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (65313)..(65313)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (80024)..(80024)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (10091)..(10091)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (102696)..(102696)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (105121)..(105121)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (107248)..(107248)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (117136)..(117136)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (119750)..(119750)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (119924)..(119924)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (12038)..(12038)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (121344)..(121344)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (122167)..(122167)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (122336)..(122336)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (131340)..(131340)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (131360)..(131360)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (139910)..(139910)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (140398)..(140398)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (142750)..(142750)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (145058)..(145058)
? OTHER INFORMATION: n equals a,t,c, or g
? FEATURE:
? NAME/KEY: misc_feature
```



	/	LOCATION:	(145171) .. (145171)	
:	/	OTHER INFORMATION:	n equals a,t,c, or g	
:	/	FEATURE:		
:	/	NAME/KEY:	misc_feature	
:	/	LOCATION:	(145942) .. (145942)	
:	/	OTHER INFORMATION:	n equals a,t,c, or g	
:	/	FEATURE:		
:	/	NAME/KEY:	misc_feature	
:	/	LOCATION:	(147197) .. (147197)	
:	/	OTHER INFORMATION:	n equals a,t,c, or g	
:	/	FEATURE:		
:	/	NAME/KEY:	misc_feature	
:	/	LOCATION:	(150841) .. (150841)	
:	/	OTHER INFORMATION:	n equals a,t,c, or g	
:	/	FEATURE:		
:	/	NAME/KEY:	misc_feature	
:	/	LOCATION:	(152500) .. (152500)	
:	/	OTHER INFORMATION:	n equals a,t,c, or g	
:	/	FEATURE:		
:	/	NAME/KEY:	misc_feature	
:	/	LOCATION:	(152530) .. (152530)	
		Alignment Scores:		
Pred. No.:	9.5e-09	Length:	1830121	
Score:	216.00	Matches:	77	
% Similarity:	45.26%	Conservative:	47	
Best Local Similarity:	28.10%	Mismatches:	118	
Query Match:	13.21%	Indels:	32	
DB:	20	Gaps:	8	
US-10-089-514-6 (1-322) x US-10-158-865-1 (1-1830121)				
Oy	2	SerglyPheProArgSerValValAlGlyIleSerglyAlaValAlGlyIleMetPheAla	21	
Dd	1370739	TCTGTATTATTCACAAATTTGTTATGTCGGCGTATAGSTTAATTAGCGCGTTATTGCC	1370798	
Oy	22	GlyLeuleuAargLualAgLySerArgThrLeuValAlAspLeuValProProGly	41	
Dd	1370799	CCTATTACGCATCTCCGCTATCCCAATTTCTATTAGAT	1370840	
Oy	42	ArGrProAspAlAcysLeuValGlYaspValThrrAlaProGlyProGluLeuAlala	61	
Dd	1370841	CGCGAAGTTGGCGCGTGCTGAAGAATTT-----	1370870	
Oy	62	LeuArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaleuValAlaVal	81	
Dd	1370871	TTAGGAGATGCTGATGTCGTATCCTTCCCTCATTAATCTCACCTTAGAACAACTT	1370930	
Oy	82	AlaProvalThrrArgLeuMetcArgProGlyAlaleuLeuAlaAspThrLeuServ--	100	
Dd	1370931	GAACGCTTMAAACCTTATTATTAAACGAAAACAATGACTTGACAGATTAAACCTTGTTAG	1370990	
Oy	101	ArgThrGlyMetAlalagluLeuAlalahisAlaProGlyValIGlnHisValGlyLeu	120	
Dd	1370991	CGTGAACCGCTAGCGAAAAATCTTGAAATTATCTGGTCTGT-----TTAGGTTTTA	1371044	
Oy	121	AenPrometPhe---AlaProAlalaglYmetThrGlyArgProValAlalaValAlaVal	139	
Dd	1371045	CATCCAATGTTTGTCGACGATATTCGAAAGTATGGCAAACAAGTGTTTGCCGTTCGAT	1371104	
Oy	140	ThrrArgAspGlyProGlyValThrrAlaleuLeuArgLeuValIguliGylgylGlyArg	159	
Dd	1371105	GGAGCTTTCTCGAACCTTATGAATGGTTACTTGACCAAAATTCAAATTTGGGGGCAAAA	1371164	
Oy	160	ProValAargLeuThrAlaGluGluHlsAspArgThrThrrAlalahrGlnAlaleuthr	179	
Dd	1371165	ATTTATCAACCAATGCTCACAGAACACATATAATATGACTTTATATCAAGCCTTCGCG	1371224	
Oy	180	HisaAlaValLeuLeuSerPheGlyLeuAlaleuAlaArgGluGlyValAspValArgAla	199	
Dd	1371225	CATTTTTCGACTTTTGGCAATGGTTTACACTTTTCCAAACGCCCATTAATCTTCGCTAAT	1371284	
Oy	200	LeuAlalahrAlaProProChiIsoGluValleuLeuAlaleuAlaArgValleu	219	

```

Db      1371285 TTATTGGCACTTTCTTCCCTATTATTCGGTTAGACTCTGGAGATGATGCTGTTATT 1371344
Qy      220 GlyGlySerProGluValTYrGlyAspIle-----GlnArgSerAsnProArgAlaAla 237
Db      1371345 GCGCAAGATGACGAGCTTTCGCGAGATATTATTATGTAGTAAAGTCAAGAAATTTAGCGGTA 1371404
Qy      238 -----SerAlaArgArgAlaLeuValGluValaLeuArgSerPheAlaAlaLeuValGly 255
Db      1371405 ATTTGAACGCTAAACAACAACTTACGACGAGCCCTTACTTCTTT----- 1371449
Qy      256 AspAspProAspArgAlaAspAlaProGlyArgAlaAspAla 269
Db      1371450 -----GAAATTAATGATGCTCAAGGTTTATTGATGCT 1371482

RESULT 12
US-10-981-687-1
; Sequence 1, Application US/10981687
; Publication No. US20050131222A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Fra
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PR186P2CID12
; CURRENT FILING DATE: 2004-11-05
; PRIOR APPLICATION NUMBER: US 10/158,865
; PRIOR FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 09/557,884
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 08/476,102
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:

```



```
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100691)..(100691)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
```

```
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (133140)..(133140)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
```

Alignment Scores:



```

Pred. No.: 9.Se-09 Length: 1830121
Score: 216.00 Matches: 47
Percent Similarity: 45.26% Conservative: 77
Best Local Similarity: 28.10% Mismatches: 118
Query Match: 13.21% Indels: 32
DB: Gaps: 8

US-10-089-514-6 (1-322) x US-10-981-687-1 (1-1830121)

QY 2 SerGlyPheProArgSerValValGlyGlySerGlyAlaValGlyGlyMetPheAla 21
DB 1370739 TCTGATATTTCACAAATGTTATGCGGGTATGGAATTAAGCGCGCTATTGCGC 1370798

QY 22 GlyLeuLeuArgGluGlySerArgThrLeuValValLeuValProProGly 41
DB 1370799 CGTTATTACGTCGATGCTGCTATCCATTCTATTATTTAGAT 1370840

QY 42 ArgProAspAlaCysLeuValGlyAspValThrAlaProGlyProGluLeuAlaAla 61
DB 1370841 CGCGAAGATTGGGGCGTGGTGAAGTAT 1370870

QY 62 LeuArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuValAlaVal 81
DB 1370871 TTAGCGAATGCTGATGCTGATGCTTCCGCTCTATTATCTCACTTAGAACAAT 1370930

QY 82 AlaProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerVal 100
DB 1370931 GAACGCTTAAACCTTATTATTAACGGAACATGCTCACTTCAAGTTTAACTCTGTTAG 1370990

QY 101 ArgThrGlyMetAlaAlaGluLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeu 120
DB 1370991 CGTGAACCGCTAGCGAATAATGCTGAAGTTCATATCTGCTGTT 1371044

QY 121 AsnProMetPhe--AlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValAl 139
DB 1371045 CATCCATGTTTGTGCGAGATATATGCAAGTATGCAAAACAGTGTGTCGCTTGAT 1371104

QY 140 ThrArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGluGlyGlyArg 159
DB 1371105 GGAAGCTTTTCTTAAACCTTATGAATGGTTATCTTGACCAATTCGAGTGGCAAA 1371164

QY 160 ProValArgLeuThrAlaGluGluHisAspArgThrThrAlaAlaThrGlnAlaLeuThr 179
DB 1371165 ATTATCAACAACCAATGCGCAACACGATCATATATGACTTATATCAAGCTTGCGC 1371224

QY 180 HisAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAla 199
DB 1371225 CATTTTTCGACTTTGCGAATGGTTATACCTTTCCAAACAGCCATTTATCTCGTAAT 1371284

QY 200 LeuAlaAlaThrAlaProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeu 219
DB 1371285 TTATGCGACTTTCTTCCCTTATTTTGGTTAGTTCGATGAGTATGCTTTATTT 1371344

QY 220 GlyGlySerProGluValTyrGlyAspIle-----GlnArgSerAsnProArgAlaAla 237
DB 1371345 GCCCAAGATGCAAGCTTTTACGAGATATTATATGATTAAGTACAGAAATTTAAGCGGTA 1371404

QY 238 -----SerAlaArgArgAlaLeuAlaGluAlaLeuArgSerPheAlaAlaLeuValGly 255
DB 1371405 ATTGAACCGTAAACAAACTTACGACGAGAGCCTTAACTTTCTT 1371449

QY 256 AspAspProAspArgAlaAspAlaProGlyArgAlaAspAla 269
DB 1371450 -----GAAATAATATGATGCTCAAGCTTTTATTGATGCT 1371482

RESULT 13
US-10-450-763-1605
; Sequence 1605, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: HySeq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US

```

```

; CURRENT APPLICATION NUMBER: US/10/450,763;
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 1605
; LENGTH: 1792
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: SIMILAR
; LOCATION: (567)..(4)
; OTHER INFORMATION: 81% homologous to Escherichia coli tRNA
; OTHER INFORMATION: nucleotidyltransferase, accession number M12788, Smith-Waterman Sc
; OTHER INFORMATION: =820.
US-10-450-763-1605

Alignment Scores:
Pred. No.: 6,66e-12 Length: 1792
Score: 213.00 Matches: 75
Percent Similarity: 39.78% Conservative: 36
Best Local Similarity: 26.88% Mismatches: 90
Query Match: 13.03% Indels: 78
DB: Gaps: 8

US-10-089-514-6 (1-322) x US-10-450-763-1605 (1-1792)

QY 49 GlyAspValThrAlaProGlyProGluLeuAlaAlaLeuArgAspAlaAspLeuVal 68
DB 289 GGTGACGTAATTTCCACCGCTCCGGA 315

QY 69 LeuLeuAlaValHisGluAspValAlaLeuValAlaValAlaProValThrArgLeuMet 88
DB 316 -----GCCCGCAGCAATTTCTGCGTGGGTGCTGATGCTGATCTGCTACATC 369

QY 89 -----ArgProGly 91
DB 370 GCTAAGAAACCCCGCCCGCGCAACCCCAACCCCAACCCATGCAACGCCGC 429

QY 91 ----- 91
DB 430 CCGCCTCCCGCCCGGAGCCCGCACCCCTATGTCATTCACCGCTTTAGGAA 489

QY 92 -----AlaLeuLeuAlaAspThrLeuSerValArgThrGly--MetAlaAlaGluLeuAla 109
DB 490 GATTGATTCTGTCATCTGGCATTCAGTGAATAATGGCCATTACAGGCCATGCTGCTG 549

QY 110 AlaHisAlaProGlyValGlnHisValGlyLeuAsnProMetPheAlaProAlaAlaGly 129
DB 550 GCGCATGATGTCGCGTG-----CTGGGGCTACACCCATGTTGCTGCGGACAGCGGT 603

QY 130 MetThrGlyArgProValAlaAlaValAlaThrArgAspGlyProGlyValThrAla 148
DB 604 AGCGTGGCAACGAAGTT-----GTGCTGCTGATGATGAGCGTAAACCGGAAGCATAC 657

QY 149 -----LeuLeuArgLeuValGluGlyGlyGlyArgProValArgLeuThrAlaGlu 166
DB 658 CAATGTTTCTGGAGCAATTCAGGCTGGGCGCTGCGCTGCATGTTATGCGCGCTGC 717

QY 167 GluHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeuLeuSerPhe 186
DB 718 GAGCAGATCAAGAAATATGCGCTTTTACGGACATCGCGCACTTTGCTATTCTTTGCTTAC 777

QY 187 GlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaValAlaAlaThrAlaProPro 206
DB 778 GGGCTGCACTGGCAGAAAGAAATGTTTCACTTGAAGCAACTTGGCGGCTCTTGGCGG 837

QY 207 ProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGlyGlySerProGluValTyr 226

```







Db	43	CTGGTGGTGGTGGTGGTGGGTTGATCGGTTGGTTCGTTTGGCCAAAGGCTTGGCTGAAGC	102
Oy	28	Gly----SerArgThrLeuVal--ValAspLeuValProProArgIyArProAspAla	45
Db	103	GCGGCTGTGGGAGAGTGTCGGGGCTGCATCTC--GACCCGAAATCGCCGAACTGGCG	159
Oy	46	CysLeuValGlyAspValThrAlaProGlyProGluLeuAlaAlaLeuArgAspAla	65
Db	160	GTGAGTGTGGGGCTGGTGGATGGTTCGGAAGACAGCTGGTGGCGCTTGGCCGGGGCGG	219
Oy	66	AspLeuValLeuLeuAlaValHisGluAspValAlaLeuLysAlaValAlaProValThr	85
Db	220	GACGGATCTCAAGTTGGCTGGTGGCCGATCTGGCCATGAAAGTGTGGCGCTTCGGCC	279
Oy	86	ArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrGlyMetAla	105
Db	280	GGCATGGATCTGGGGGAGCGAGCTCTGACAGGATGTGGCGAGCCCAAG--GGCAATGTG	336
Oy	106	AlaGluLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeuAsnProMetPheAla	125
Db	337	GTGGCTGGGGGAGACCGAAGCGTTTCGGC-----GGCATGGCGGGCGCTTTCGTG	384
Oy	126	ProAlaAlaGlyMetThrGlyArgProValAlaAlaVal-----	138
Db	385	CCGGGCGCATCCGATTGGCCGGTCCGAGCAGAGCGGGGTGGAGCTTCCAATCAGAGCTG	444
Oy	139	-----ValThrArgAspGlyProGlyValThrAla	148
Db	445	TTCCGTCGTGATAAAGTATCTGACCCCGGTGGAGACAGCGATCCGGCGGCTGGCC	504
Oy	149	LeuLeu---ArgLeuValGluGlyGlyGlyArgProValArgLeuThrAlaGluGlu	167
Db	505	GTGGTGGATCGCTTGTGGCGCTGAGCTGGGGCGCCGAGCTGAGCAGATGCAAGTCAAGCT	564
Oy	168	HisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeuLeuSerPheGly	187
Db	555	CACGATGAAGTCTGGCCCGCAGACAGCTATCTCCGCAC-----CTGCTGGCTTCGCT	618
Oy	188	Leu-----AlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaAlaThrAlaPro	205
Db	619	CTGGTGGATTCACGTGGCCAAAGCGCAATGAAGACTTGAGATCTTCGTTACGCTGGGGC	678
Oy	206	ProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGlySerProGluVal	225
Db	679	GATTTCCTGAT-----TTCACAAAGATGCGCCGAAAGCGACCCGCTGCATG	723
Oy	226	TyrGlyAspIleGlnArgSerAsnProArgAlaAlaSerAlaArgArgAlaLeuAlaGlu	245
Db	724	TGGCATGACATTTTCTCGCCAC-----CGCAGAGCTGTCTGGCCG	765
Oy	246	AlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArgAlaAspAlaProGly	265
Db	766	ACACTCGATACATTTTCGAGCAGACCTCGACGCTTGGCGGACCGGATCATGCA-----	819
Oy	266	ArgAlaAspAlaProGlyHisProGlyGlyCysAspGlyAlaGlyAsnLeuAspGlyVal	285
Db	820	-----GGGATGGGACAC-----CAATTGTGGCGGTC	846
Oy	286	PheGlyGluLeuArg	290
Db	847	TTTACGCGGCGCCCG	861

Search completed: October 6, 2005, 02:37:03  
Job time : 3498.63 secs



This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: October 5, 2005, 09:03:04 ; Search time 3522.58 Seconds  
(without alignments)  
3479.464 Million cell updates/sec

Title: US-10-089-514-6  
Perfect score: 1635  
Sequence: 1 MSGPRSVVVGSGAVGWMF.....HCELFRLTHTDDEGKDR 322

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues  
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+.p2n.model -DEV=xlh  
-Q/cn2.1/USPTO.spool.h/US10089514/runat\_04102005.105744.8043/app.query.fasta.1.1621  
-DB=EST -OPMT=fastrp -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=bl0sum62 -TRANS=human40.cdi -LIST=45  
-DOCAIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=prc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10089514\_@CN2\_1\_1.6050\_@runat\_04102005.105744.8043 -NCPU=6 -ICPU=3  
-NO\_MMMP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONELIG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.\*  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hic.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_gsa1.\*  
9: gb\_gsa2.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	202.5	12.4	597	9	AG020402 Homo sapi
2	196.5	12.0	677	9	AG020403 Homo sapi
3	162.5	9.9	851	7	CF823705 EST701087
4	162.5	9.9	875	7	CO022117 EST818210
5	154.5	9.4	848	7	CO003446 EST791781
6	154.5	9.4	900	7	CF823364 EST700746
7	152.5	9.3	973	7	CF881643 tr1C028xo
8	146.5	9.0	821	7	CF682058 CCAHL63TR
9	146.5	9.0	830	7	CF702754 CCAHK69TR

10	146	8.9	859	7	CF819790	CF819790 EST697172
C 11	145.5	8.9	1407	9	CL959323	AG059323 OaIFCC002
C 12	143.5	8.8	969	9	AG061086	AG061086 Pan trogl
C 13	141	8.6	1086	9	AY408045	AY408045 Homo sapi
C 14	141	8.6	1684	3	CR595067	CR595067 full-1eng
C 15	140.5	8.6	1145	9	AG071167	AG071167 Pan trogl
16	139	8.5	820	6	CD616495	CD616495 56075281H
17	138.5	8.5	751	7	CF882803	CF882803 tr1C087xe
18	138.5	8.5	805	6	CB909122	CB909122 tr1C087xe
C 19	138.5	8.5	984	5	BQ053468	BQ053468 AGENCOURT
C 20	138.5	8.5	991	5	BM220915	BM220915 AGENCOURT
21	137.5	8.4	713	7	CF676551	CF676551 CCAH141TR
22	137.5	8.4	747	7	CF721128	CF721128 CCAH142TR
C 23	136.5	8.3	2176	3	CR603382	CR603382 full-1eng
C 24	134.5	8.2	993	3	CL958449	CL958449 OaIFCC001
25	133.5	8.2	760	6	CB902069	CB902069 tr1C028xo
26	133.5	8.2	760	7	CF871839	CF871839 tr1C028xo
27	132.5	8.1	1098	7	CK210037	CK210037 FGAS02182
C 28	132.5	8.1	1122	5	BM928388	BM928388 AGENCOURT
C 29	131.5	8.0	697	9	CC727568	CC727568 OaIFCC004
C 30	131	8.0	865	5	BR372494	BR372494 BX372494
31	130.5	8.0	1309	3	CR723473	CR723473 Tetradon
32	130	8.0	812	4	BG648802	BG648802 EST510421
C 33	129.5	7.9	1116	8	BZ569478	BZ569478 PAC62-164
34	129	7.9	1207	5	BQ645487	BQ645487 AGENCOURT
35	129	7.9	3405	9	CL960595	CL960595 OaIFCC004
C 36	128.5	7.9	1215	9	AG092472	AG092472 Pan trogl
37	128	7.8	843	7	CF210859	CF210859 CMB20007
38	128	7.8	904	4	BG757997	BG757997 602715041
C 39	128	7.8	1195	9	AG082926	AG082926 Pan trogl
C 40	128	7.8	1758	9	CL965380	CL965380 OaIFCC012
C 41	127.5	7.8	724	2	AM661060	AM661060 833001B03
C 42	127.5	7.8	1056	9	CNS04UNT	AL233782 Tetradon
C 43	127.5	7.8	1062	4	BM544107	BM544107 AGENCOURT
C 44	127.5	7.8	1402	9	CL507797	CL507797 SAIL_788
C 45	126.5	7.7	2328	9	CL982230	CL982230 OaIFSC047

## ALIGNMENTS

RESULT 1  
AG020402/c  
LOCUS AG020402 597 bp DNA linear GSS 09-JAN-2003  
DEFINITION Homo sapiens genomic DNA, 21q region, clone: B53P19\_BB042(Fw).abl,  
genomic survey sequence.

ACCESSION AG020402  
VERSION AG020402.1 GI:6046346  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1  
Hukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Hattori,M., Fujiyama,A., Ishii,K., Toyoda,A., Taylor,T.,  
Park,H.-S., Yada,T., Watanabe,H. and Sakaki,Y.

TITLE Homo sapiens genomic DNA, chromosome 21q  
JOURNAL Published Only in Database (1999)

REFERENCE 2 (bases 1 to 597)  
Hattori,M., Fujiyama,A., Ishii,K., Toyoda,A., Taylor,T.,  
Park,H.-S., Yada,T., Watanabe,H. and Sakaki,Y.

TITLE Direct Submision  
JOURNAL Submitted (16-OCT-1999) Maehara Hattori, RIKEN Genomic Sciences  
Center(GSC) c/o Kitasato University, 1-15-1 Kitasato, Sagamihara,  
Kanagawa 228-8555, Japan (E-mail:hattori@hg.ims.u-tokyo.ac.jp,  
Tel:81-42-778-9923, Fax:81-42-778-9924)

FEATURES  
source 1. 597  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q"  
/clone="B53P19\_BB042(Fw).abl"



ORIGIN

Alignment Scores: 6.81e-08 Length: 597

Pred. No.: 202.50 Matches: 65

Score: 49.30% Conservative: 41

Percent Similarity: 30.23% Mismatches: 84

Best Local Similarity: 12.39% Indels: 25

Query Match: 9 Gaps: 7

DB: 9

US-10-089-514-6 (1-322) x AG020403 (1-597)

QY 8 ValValaValGlyGlySerGlyAlaValGlyGlyMetPheAlaGlyLeuAlaArgGluAla 27

DB 595 GTTATCGTCGGCGCGTCGCGATGGAGCGCCTGTCGAGAAGATGCTGACCTCTCG 536

QY 28 GlySerArgThrLeuValAlaAspLeuValProProProGlyArgProAspAlaCysLeu 47

DB 535 GGTATTCAGGTGGCGGATTCGGAG-----CAACATGACCTGGATCGA 494

QY 48 ValGlyAspValThrAlaProGlyProGluLeuAlaAlaLeuArgAspAlaAspLeu 67

DB 493 GCGGCTGATATGTGTGCG-----GATGCCGGAATG 464

QY 68 ValLeuLeuAlaValAlaHisGlyAspValAlaLeuLysAlaValAlaProValThrArgLeu 87

DB 463 GTGATGTGATGATGTCGCAATCCAGTTACTGACCAAGTTATGCGCAATTACCGCCTTTA 404

QY 88 MetArgProGlyAlaLeuAlaAspThrLeuSerValArgThrGly---MetAlaAla 106

DB 403 CCGAAA---GATTGTATTCTGTCGATCTGCGATCGATGTAATAATGGCCATTACAGGCC 347

QY 107 GluLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeuAsnProMetPheAlaPro 126

DB 346 ATGCTGTGGCGCATGATGTCGGTG-----CTGGGCTACACCGGATGTTCCGTCG 293

QY 127 AlaAlaGlyMetThrGlyArgProValAlaAlaValAlaThrArgAspGlyProGlyVal 146

DB 292 GACAGCGGTAGCCTGGCAAGCAAGTT-----GTGGTCGTGTGATGAGCACTGAACCG 239

QY 147 ThrAla-----LeuLeuArgLeuValGluGlyGlyGlyArgProValArgLeu 163

DB 238 GAAGCATACCAATGTTTCTGGAGCAATTCAGGTCGGGGCGCTGGCTGCATCGATTT 179

QY 164 ThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeu 183

DB 178 AGCGCGGTGAGCAGCATGAGATATGGCGTTATTTCAGGCACTGGCCACTTGTCTACT 119

QY 184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaAlaThr 203

DB 118 TTTGCTTACGGGCTGCACCTGGCAGAAATGTTTCAGCTTGAGCAACTTTCGGCGCTC 59

QY 204 AlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgVal 218

DB 58 TCTTCGCCGATTTACCGCCTTGAGCTGGCGATGTCGGGAGACTG 14

RESULT 2

AG020403 677 bp DNA linear GSS 09-JAN-2003

LOCUS Homo sapiens genomic DNA, 21q region, clone: B53P19\_BB042 (Rev. a),

DEFINITION genomic survey sequence.

ACCESSION AG020403

VERSION AG020403.1 GI:6046347

KEYWORDS GSS.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 Hattori,M., Fujiyama,A., Ishii,K., Toyoda,A., Taylor,T., Park,H.-S., Yada,T., Watanabe,H. and Sakaki,Y. Mammalia; Eutheria; Primates; Catarrhini; Homiindae; Homo.

AUTHORS Hattori,M., Fujiyama,A., Ishii,K., Toyoda,A., Taylor,T., Park,H.-S., Yada,T., Watanabe,H. and Sakaki,Y.

TITLE Homo sapiens genomic DNA, chromosome 21q

JOURNAL Published Only in Database (1999)

REFERENCE 2 (bases 1 to 677)

AUTHORS Hattori,M., Fujiyama,A., Ishii,K., Toyoda,A., Taylor,T., Park,H.-S., Yada,T., Watanabe,H. and Sakaki,Y.

TITLE Direct Submission

JOURNAL Submitted (16-OCT-1999) Maehira Hattori, RIKEN Genomic Sciences Center (GSC) c/o Kitasato University, 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan (E-mail:hattori@hgsc.ims.u-tokyo.ac.jp, Tel:81-42-778-9923, Fax:81-42-778-9924)

FEATURES

Location/Qualifiers

1..677

source

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="21"

/map="21q"

/clone="B53P19\_BB042 (Rev. a)"

ORIGIN

Alignment Scores: 2.71e-07 Length: 677

Pred. No.: 196.50 Matches: 64

Score: 49.06% Conservative: 40

Percent Similarity: 30.19% Mismatches: 83

Best Local Similarity: 12.02% Indels: 25

Query Match: 9 Gaps: 7

DB: 9

US-10-089-514-6 (1-322) x AG020403 (1-677)

QY 8 ValValaValGlyGlySerGlyAlaValGlyGlyMetPheAlaGlyLeuAlaArgGluAla 27

DB 101 GTTATCGTCGGCGCGTCGCGATGGAGCGCCTGTCGAGAAGATGCTGACCTCTCG 160

QY 28 GlySerArgThrLeuValAlaAspLeuValProProProGlyArgProAspAlaCysLeu 47

DB 161 GGTATTCAGGTGGCGGATTCGGAG-----CAACATGACCTGGATCGA 202

QY 48 ValGlyAspValThrAlaProGlyProGluLeuAlaAlaLeuArgAspAlaAspLeu 67

DB 203 GCGGCTGATATGTGTGCG-----GATGCCGGAATG 232

QY 68 ValLeuLeuAlaValAlaHisGlyAspValAlaLeuLysAlaValAlaProValThrArgLeu 87

DB 233 GTGATGTGATGTCGCAATCCAGTTACTGACCAAGTTATGCGCAATTACCGCCTTTA 292

QY 88 MetArgProGlyAlaLeuAlaAspThrLeuSerValArgThrGly---MetAlaAla 106

DB 293 CCGAAA---GATTGTATTCTGTCGATCTGCGATCGATGTAATAATGGCCATTACAGGCC 349

QY 107 GluLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeuAsnProMetPheAlaPro 126

DB 350 ATGCTGTGGCGCATGATGTCGGTG-----CTGGGCTACACCGGATGTTCCGTCG 403

QY 127 AlaAlaGlyMetThrGlyArgProValAlaAlaValAlaThrArgAspGlyProGlyVal 146

DB 404 GACAGCGGTAGCCTGGCAAGCAAGTT-----GTGGTCGTGTGATGAGCACTGAACCG 457

QY 147 ThrAla-----LeuLeuArgLeuValGluGlyGlyGlyArgProValArgLeu 163

DB 458 GAAGCATACCAATGTTTCTGGAGCAATTCAGGTCGGGGCGCTGGCTGCATCGATTT 517

QY 164 ThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeu 183

DB 518 AGCGCGGTGAGCAGCATGAGATATGGCGTTATTTCAGGCACTGGCCACTTGTCTACT 577

QY 184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaAlaThr 203

DB 578 TTTGCTTACGGGCTGCACCTGGCAGAAATGTTTCAGCTTGAGCAACTTTCGGCGCTC 637

QY 204 AlaProProProHisGlnValLeuLeuAlaLeuAlaLeu 215

DB 638 TCTTCGCCGATTTACCGCCTTGAGCTGGCGATGTC 673

RESULT 3



CF823705	CF823705	551 bp	mRNA	linear	EST 01-APR-2004
LOCUS	EST701087	Coccidioides posadasii saprobic phase cDNA library, 2 to 4 kb	Coccidioides posadasii cDNA clone CIDAN17 5' end, mRNA sequence.		
DEFINITION					
ACCESSION	CF823705				
VERSION	CF823705.1				
KEYWORDS	GI:45929762				
SOURCE	EST.				
ORGANISM	Coccidioides posadasii				
REFERENCE	Coccidioides posadasii				
AUTHORS	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Onygenales; mitosporic Onygenales; Coccidioides.				
TITLE	1 (bases 1 to 551)				
JOURNAL	Gardner, M. J. and Cole, G. T.				
COMMENT	Analysis of gene expression in Coccidioides posadasii mycelia and spherules via expressed sequence tags unpublished (2003)				
OTHER ESTs	Other ESTs: EST701086				
Contact:	Gardner MJ				
The Institute for Genomic Research					
9712 Medical Center Drive, Rockville, MD 20850, USA					
Tel: 301 838 3519					
Fax: 301 838 0208					
Email: gardner@igr.org					
Seq primer: M13 Reverse					
Location/Qualifiers					
1..851					
/organism="Coccidioides posadasii"					
/mol_type="mRNA"					
/strain="C735"					
/db_xref="taxon:199306"					
/clone="CIDAN17"					
/dev_stage="saprobic phase (mycelia)"					
/lab_host="E. coli DH10B, T1 phage resistant"					
/clone_lib="Coccidioides posadasii saprobic phase cDNA library, 2 to 4 kb"					
/note="Vector: pExpress 1; Site_1: Not I; Site_2: Eco RV; Coccidioides posadasii saprobic phase cDNA library, size fractionated cDNA 2 to 4 kb"					
ALIGNMENT SCORES:					
Pred. No.:	0.000364	Length:	851		
Score:	162.50	Matches:	60		
Percent Similarity:	44.72%	Conservative:	29		
Best Local Similarity:	30.15%	Mismatches:	85		
Query Match:	9.94%	Indels:	25		
DB:	7	Gaps:	8		
US-10-089-514-6 (1-322) x CF823705 (1-851)					
OY	5	Proatgservalaivaalgaiglysergylalavalaglygmetphealaglyleu	24		
DB	260	CCAAAGATCATTGGGATCATTTGGGATGGGTATATGGGAAGATGTATGCTCAGCGCTTG	319		
OY	25	Argglunlaglyserargthrlleuvalalaspheuvalproprogllyarproasp	44		
DB	320	AGCGAAGGGGT-----TGGAGCATTAATGCTGCGCAACAACCTACCAATTATGAC	370		
OY	45	Alacyleuvalglaspvalthr-----Alaproglyproglu	58		
DB	371	ACACTGAGGCAAGATTCATCCATCAATATCAATATATTAACCAACGACATTGG	430		
OY	59	Alalalalaleuargaspalaspheuleuvalaleuvalahlsglaspvalalaleu	78		
DB	431	GTCCTC-----AGAGTTAGGACCTACTATCTACAGTGTAGAGGCAAGCCATCAAC	484		
OY	79	lysalaalalalaprovalthrargleuemetarproglialaleuleuAlaspThrlleu	98		
DB	485	AAATTTGGGGGAGTATGCGCCATCAACCAAAAGTTGCGCGATTGTGTCGCAACT	544		
OY	99	ServalargthrglymetalalaglualeuAlala-----Hlealaprogly--	114		

Db 545 TCCTGCAAA-----GCCCCAGAGCTGGCTCTTTTGAAGAACATCTTCTCGGGAC 595

Qy 115 ValGlnHisValIglyLeuAsnProMetEphelAProAlaAlaIglyMetThrGlyArgPro 134  
 Db GTGGAGATTATTTGGTTCATTCCCTCATATGAGCCGAACGTGAACCCCTAAAGAGACCG 655

Qy 135 ValAlaAlaValThrArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGlu 154  
 Db 656 TTGGCTTGATTTCAGCACAGA-----GCCCTGACGAGACCTTCAGGTTTGTGAA 706

Qy 155 -----GlyGlyGlyArgProValArgLeuThrAlaGluGluHisAspArg 170  
 Db 707 AGTGGTTGGCTCTTCGATTCCAATACGTTACTACCGGGAGAGCATGATCGA 766

Qy 171 ThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeuLeuSerPheGlyLeuAla 189  
 Db 767 ATCACAGGAGCACTCAGCGCCGTGACGATCTCTGCATTTTGAATGAGGAACCTGCG 823

RESULT 4  
 C0022117  
 LOCUS C0022117  
 DEFINITION ESF818210 Coccidioides posadasii saprobic phase cDNA library, greater than 4kb Coccidioides posadasii cDNA clone CIBCR06 5' end, mRNA sequence.  
 C0022117  
 EST. C0022117.1 GI:48546838  
 SOURCE Coccidioides posadasii  
 ORGANISM Coccidioides posadasii  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Onygenales; mitosporic Onygenales; Coccidioides.  
 1 (bases 1 to 875)  
 Gardner,M.J. and Cole,G.T.  
 Analysis of gene expression in Coccidioides posadasii mycelia and spherules via expressed sequence tags  
 Unpublished (2003)  
 Other\_ESTs: ESF818209  
 JOURNAL Contact: Gardner MJ  
 COMMENT The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301 838 3519  
 Fax: 301 838 0208  
 Email: gardner@ligr.org  
 Seq primer: M13 Reverse.

FEATURES  
 source  
 Location/Qualifiers  
 1..875  
 /organism="Coccidioides posadasii"  
 /mol\_type="mRNA"  
 /strain="C735"  
 /db\_xref="taxon:199306"  
 /clone="CIBCR06"  
 /dev\_stage="saprobic phase (mycelia)"  
 /lab\_host="E. coli DH10b, TI phage resistant"  
 /clone\_idb="Coccidioides posadasii saprobic phase cDNA library, greater than 4kb"  
 /note="Vector: pExpress 1; Site 1: Not I; Site 2: Eco RV; Coccidioides posadasii saprobic phase cDNA library, size fractionated cDNA > 4 kb"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.000377 Length: 875  
 Score: 162.50 Matches: 60  
 Percent Similarity: 44.72% Conservative: 29  
 Best Local Similarity: 30.15% Mismatches: 85  
 Query Match: 9.94% Indels: 25  
 Gaps: 8

US-10-089-514-6 (1-322) x C0022117 (1-875)

Qy 5 ProArgSerValValValIglyIserGlyAlaValGlyIleMetPheAlaGlyLeu 24  
 Db 295 CCAAGATCCATTGGGATTCGGATGGGATATGGAAGATGTGTGCTCAGCCGCTT 354



```
QY 25 ArgGluAlaGlySerArgThrLeuValValAspLeuValProProGlyArgProAsp 44
    |||
    |||
    |||
Db 355 AGCCAGCGGGT-----TCGACATTAATGCTGGCAGCAAACTACCAATTATGAC 405
    |||
    |||
    |||
QY 45 AlaGlyLeuValGlyAspValThr-----AlaProGlyProGlyLeu 58
    |||
    |||
    |||
Db 406 ACACAGGAGCAAGAAATTCATATCCATACAAATATCATATATTCACCAACGACATTGG 465
    |||
    |||
    |||
QY 59 AlaAlaAlaLeuArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeu 78
    |||
    |||
    |||
Db 466 GTCCTC-----AGAGTTAGCGACTACATTATCTACAGTTAGAGGACAGAACCATCAAC 519
    |||
    |||
    |||
QY 79 LysAlaValAlaProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeu 98
    |||
    |||
    |||
Db 520 AAAATGTGTGGCGGAGATGGCCCATCAACCAAGTTGGCGGATGTGGTGTCAACT 579
    |||
    |||
    |||
QY 99 SerValArgThrGlyMetAlaAlaGluLeuAla-----HisAlaProGly--- 114
    |||
    |||
    |||
Db 580 TCCTGCAAA-----GCCCGAGAGCTGGCTTTGAGAACCATTCTTCTGGGGGAC 630
    |||
    |||
    |||
QY 115 ValGlnHisValGlyLeuAsnProMetPheAlaProAlaAlaGlyMetThrGlyArgPro 134
    |||
    |||
    |||
Db 631 GTGGAAATTAATTCGCTTATTCCTTACATGACCGACGACCTTAAAGACAGCCG 690
    |||
    |||
    |||
QY 135 ValAlaAlaValAlaThrArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGlu 154
    |||
    |||
    |||
Db 691 TTGGTCTGATTACAGACAGA-----GCCTGTGACGAGCCTCAAGTTGTGTA 741
    |||
    |||
    |||
QY 155 -----GlyGlyGlyGlyArgProValArgLeuThrAlaGluLuhisAspArg 170
    |||
    |||
    |||
Db 742 AGTGTGTTTGGCTCTTTCGAATCAAAATAGCTTACTCACCGGAGAGGCTGATCA 801
    |||
    |||
    |||
QY 171 ThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeuLeuSerPheGlyLeuAla 189
    |||
    |||
    |||
Db 802 ATCACAGCGGACACTAGGCGCGGACGACATGCTGCAATTTTGATGGAAGCTGGC 858
    |||
    |||
    |||
RESULT 5
LOCUS CO003446 848 bp mRNA linear EST 09-JUN-2004
DEFINITION EST791781 Coccidioides posadasii spherule cDNA library, 0.4 to 2.3
kb Coccidioides posadasii cDNA clone C1EAD69 5' end, mRNA sequence.
ACCESSION CO003446
VERSION CO003446.1 GI:48510335
KEYWORDS EST.
SOURCE Coccidioides posadasii
ORGANISM Coccidioides posadasii
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; Mucosporic Onygenales; Coccidioides.
REFERENCE
AUTHORS Gardner,M.J. and Cole,G.T.
TITLE Analysis of gene expression in Coccidioides posadasii mycelia and
spherules via expressed sequence tags
JOURNAL Unpublished (2003)
COMMENT Other_ESTs: EST791780
Contact: Gardner MJ
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@tigr.org
Seg Primer: M13 Reverse.
Location/Qualifiers
1..848
/organism="Coccidioides posadasii"
/mol_type="mRNA"
/strain="C735"
/db_xref="taxon:1993306"
/clone="C1EAD69"
/dev_stage="spherules"
/lab_host="E. coli DH10B, T1 phage resistant"
/clone_11b="Coccidioides posadasii spherule cDNA library,
0.4 to 2.3 kb"
```

## ORIGIN

/note="Vector: pExpress 1; Site\_1: Not I; Site\_2: Eco RV;  
Coccidioides posadasii spherule cDNA library, 0.4 to 2.3  
kb"

Alignment Scores:

Pred. No.:	Length:
0.00185	848
Score: 154.50	Matches: 66
Percent Similarity: 41.70%	Conservative: 32
Best Local Similarity: 28.09%	Mismatches: 94
Query Match: 9.45%	Indels: 44
DB: 7	Gaps: 10

US-10-089-514-6 (1-322) x CO003446 (1-848)

```
QY 63 ArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuLysAlaValAla 82
    |||
    |||
    |||
Db 62 AGAGTTAGCGGACTTACTTATCTACAGTGTAGAGGACGAGACCATCAACAAAATTGTGGCG 121
    |||
    |||
    |||
QY 83 ProValThrArgLeuMetArgProGlyAlaLeuAlaAspThrLeuSerValArgThr 102
    |||
    |||
    |||
Db 122 GAGTATGGGCCCATCAACCAAGTTGGCGGATTTGGTGTGCTCAAACTTCTGGCAAA--- 178
    |||
    |||
    |||
QY 103 GlyMetAlaAlaGlyLeuAlaAla-----HisAlaProGly---ValGlnHisVal 118
    |||
    |||
    |||
Db 179 -----GCCCGAGAGCTGGCTCTTTTGAACAACATCTTCTGGGAGCGTGAGATTATT 232
    |||
    |||
    |||
QY 119 GlyLeuAsnProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaValAl 138
    |||
    |||
    |||
Db 233 TCGGTTCAATCCCTTACATGACCGAAGCTGAACCTTAAAGACAGCGTTGGTCTTGATT 292
    |||
    |||
    |||
QY 139 ValThrArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGlu----- 154
    |||
    |||
    |||
Db 293 CAGCACAG-----GCCTGTGACGAGACCTCAGGTTGTGTAAGTGTGGC 343
    |||
    |||
    |||
QY 155 GlyGlyGlyGlyArgProValArgLeuThrAlaGluLuhisAspArgThrThrAlaAla 174
    |||
    |||
    |||
Db 344 TCTTTCGAATCCAAATACGTTTACTCACCAGGAGAGAGCATGATCATCAGCGGAC 403
    |||
    |||
    |||
QY 175 ThrGlnAlaLeuThrHisAlaValLeuLeuSerPheGlyLeuAla----- 189
    |||
    |||
    |||
Db 404 ACTCAGGCCGTGACGAGTGTGCTCATTTTGAGTATGGGAACGTCGGCGACCACTAATAC 463
    |||
    |||
    |||
QY 190 -----LeuAlaArgLeu-----GlyValAspValArgAlaLeuAlaAla 202
    |||
    |||
    |||
Db 464 CAATCCCGTGGAGATATCTGATATGCGGTGAATTGA----- 505
    |||
    |||
    |||
QY 203 ThrAlaProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGlyGlySer 222
    |||
    |||
    |||
Db 506 -----AACGTCAAAATTAATATCATCACTTACGCAATCTCAATTAAG 547
    |||
    |||
    |||
QY 223 ProGluValIlyrGlyAspIleGlnArgSerAsnProArgAlaAlaSerIlaArgArgAla 242
    |||
    |||
    |||
Db 548 TGGCATGTCTATGCTGAGCTGCGATTTGAACCTCGCTCCAAAAGACAGATCCGCTCAG 607
    |||
    |||
    |||
QY 243 LeuAlaGluAlaLeuArgSer---PheAlaAlaLeuValGlyAspAspProAspArgAla 261
    |||
    |||
    |||
Db 608 TATGCAGATTAAGTACGACCAACTATTTCAATTGATGTTAGCGCGCATGTGATGA-GCT 666
    |||
    |||
    |||
QY 262 AspAlaProGlyArgAlaAspAlaAspAlaAspGlyHisProGlyGlyCys 276
    |||
    |||
    |||
Db 667 -----TAAGCGCAGAGTCAAGGCTCCAGCGGCGCTGT 699
    |||
    |||
    |||
RESULT 5
LOCUS CF823364 900 bp mRNA linear EST 01-APR-2004
DEFINITION EST700746 Coccidioides posadasii saprobic phase cDNA library, 2 to
4 kb Coccidioides posadasii cDNA clone CID410 5' end, mRNA
sequence.
ACCESSION CF823364
VERSION CF823364.1 GI:45929421
KEYWORDS EST.
SOURCE Coccidioides posadasii
```



ORGANISM Coccidioides posadasii  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Onygenales; mitosporic Onygenales; Coccidioides.

REFERENCE 1 (bases 1 to 900)  
Gardner M.J. and Cole G.T.  
Analysis of gene expression in Coccidioides posadasii mycelia and spherules via expressed sequence tags  
Unpublished (2003)  
Other ESTs: EST700745

JOURNAL Contact: Gardner MJ  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301 838 3519  
Fax: 301 838 0208  
Email: gardner@igf.org  
Seq primer: M13 Reverse.

COMMENT Location/Qualifiers  
1..900  
/organism="Coccidioides posadasii"  
/mol\_type="mRNA"  
/strain="C735"  
/db\_xref="taxon:199306"  
/clone="CIDAL10"  
/dev\_stage="saprobiic phase ('mycelial')"  
/lab\_host="E. coli DH10B, T1 phage resistant"  
/clone\_lib="Coccidioides posadasii saprobiic phase cDNA library, 2 to 4 kb"  
/note="Vector: pExpress 1; Site\_1: Not I; Site\_2: Eco RV; Coccidioides posadasii saprobiic phase cDNA library, size fractionated cDNA 2 to 4 kb"

ORIGIN

Alignment Scores:  
Pred. No.: 0.00199 Length: 900  
Score: 154.50 Matches: 66  
Percent Similarity: 41.70% Conservative: 32  
Best Local Similarity: 28.09% Mismatches: 94  
Query Match: 9.45% Indels: 44  
DB: 7 Gaps: 10

US-10-089-514-6 (1-322) x CF823364 (1-900)

QY 63 ArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuValAlaValAla 82  
DB 125 AGAGTTAGCGCATTAATCTACAGTGTAGAGCCAAACCATCAACAAATTTGGCG 164  
QY 83 ProValThrArgLeuMetArgProGlyValAlaLeuLeuAlaAspThrLeuSerValArgThr 102  
DB 185 GAGTATGGCCCATCAACCAAGTTGGCGCATTTGTTGCAACTTCTGCAAA--- 241  
QY 103 GlyMetAlaAlaGluLeuAlaAla-----HisAlaProGly---ValGlnHisVal 118  
DB 242 -----GCCCAAGAGCTGGCTTTGAGAACCATTTCTCGGGAGCGTGAATATT 295  
QY 119 GlyLeuAsnProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaVal 138  
DB 296 TCGGTTCAATTCCTCATCGACCAAGCAAGTAAAGCAGCGCTGTTGATT 355  
QY 139 ValThrArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGlu----- 154  
DB 356 CAGCACAGA-----GCCTCTGACGAGAGCCTCAGGTTTGTGAAGTGTTCGCG 406  
QY 155 GlyGlyGlyArgProValArgLeuThrAlaGluGlnHisAspArgThrThrAlaAla 174  
DB 407 TCTTTGGAATCCAAATATCGTTTACCTCACCAGGGAGAGGATGTAATCACAGCGGAC 466  
QY 175 ThrGlnAlaLeuThrHisAlaValLeuLeuSerPheGlyLeuAla----- 189  
DB 467 ACTCAGGCGGTGACGCTGCTGATTTTGAATGATGGAAGCTGGGCAAGCCAAATAC 526  
QY 190 -----LeuAlaArgLeu-----GlyValAspValArgAlaLeuAlaAla 202  
DB 527 CAATTCCCGTGGAGATATCTCGATATGTGGGTGAATGAA----- 568

QY 203 ThrAlaProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGlySer 222  
DB 569 -----AACGTCAAAATTAATATACATTCAGCATTCATTCGAATAG 610  
QY 223 ProGluValTyrGlyAspIleGlnArgSerAsnProAlaAlaAlaSerAlaArgAla 242  
DB 611 TGGCATGTCTATCTGAGCTGCGATATGAACCCGTGCGCAAAAGACAGATCCGTAC 670  
QY 243 LeuAlaGluAlaLeuArgSer---PheAlaAlaLeuValGlyAspAspProAlaArgAla 261  
DB 671 TATGCAAGATCAGTGCACCAACTATTCAAATTGATGTAGCGCCCATCGTATGA-GCT 729  
QY 262 AspAlaProGlyArgAlaAspAlaProGlyHisProGlyGlyCys 276  
DB 730 -----TAAAGGAGAGTCAAGGCTGCCAGGCGGCTGT 762

RESULT 7  
CF881643 973 bp mRNA linear EST 31-OCT-2003  
LOCUS trico28xo14.b1 T. reesei mycelial culture, version 6 October 2003  
DEFINITION Hypocrea jecorina cDNA clone trico28xo14, mRNA sequence.  
ACCESSION CF881643  
VERSION CF881643.1 GI:38136325  
KEYWORDS EST.  
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)  
ORGANISM Hypocrea jecorina  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
1 (bases 1 to 973)  
Tennissen, P.-J.M., van Solingen, P., Dankmeyer, L., Mitchell, T.K., Ward, M. and Dean, R.A.  
Characterization of the protein processing and secretion pathways in a comprehensive set of expressed sequence tags from Trichoderma reesei  
FEMS Microbiol. Lett. 230 (2), 275-282 (2004)  
CONTACT: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph\_dean@ncsu.edu  
Seq primer: LT-F1 primer.  
Location/Qualifiers  
1..973  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="Qm6a"  
/db\_xref="taxon:51453"  
/clone="trico28xo14"  
/dev\_stage="mycelial"  
/clone\_lib="T. reesei mycelial culture, version 6 October 2003"  
/note="Vector: pREP3Y; Site\_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Alignment Scores:  
Pred. No.: 0.0033 Length: 973  
Score: 152.50 Matches: 59  
Percent Similarity: 42.86% Conservative: 25  
Best Local Similarity: 30.10% Mismatches: 85  
Query Match: 9.33% Indels: 27  
DB: 7 Gaps: 6

US-10-089-514-6 (1-322) x CF881643 (1-973)

QY 12 GlySerGlyAlaValGlyGlyMetPheAlaGlyLeuLeuArgGluAlaGlySerArgThr 31  
DB 169 GGCATGGGCGCATGTGGCGGATGTACGCCGAAGACTGTCCGAGACTGCTGGAGATTC 228







ACCESSION CF702754  
 VERSION CF702754.1 GI:41556913  
 KEYWORDS EST.  
 SOURCE Cryptococcus neoformans var. neoformans (Filobasidiella neoformans var. neoformans)  
 ORGANISM Cryptococcus neoformans var. neoformans  
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes; Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.  
 REFERENCE 1 (bases 1 to 830)  
 Loftus, B.  
 End sequencing of clones from a full length enriched, normalized JEC21 cDNA library  
 JOURNAL Unpublished (2003)  
 COMMENT Other ESTs: CCANR69TF  
 Contact: Brendan Loftus  
 TIGR  
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-3543  
 Fax: 301-838-0208  
 Email: crypt@tigr.org  
 Seq primer: TR.  
 FEATURES  
 source  
 1. 830  
 Location/Qualifiers  
 /organism="Cryptococcus neoformans var. neoformans"  
 /mol\_type="mRNA"  
 /strain="JEC21"  
 /db\_xref="taxon:40410"  
 /clone="CCANR69"  
 /note="Vector: pCMVSPORT6; Site\_1: NotI EcoRV; The full length, normalized library was prepared from a variety of conditions using RNA provided by Joseph Hellman and Jennifer Lodge"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.00915 Length: 830  
 Score: 146.50 Matches: 72  
 Percent Similarity: 43.55% Conservative: 36  
 Best Local Similarity: 29.03% Mismatches: 110  
 Query Match: 8.96% Indels: 31  
 DB: 7 Gaps: 11  
 US-10-089-514-6 (1-322) x CF702754 (1-830)  
 QY 9 ValValIGlY-----GlySerGIYAlaValIGlYMePhaAlaGlyLeuLeuArgGlu 26  
 Db 94 GTGCTAGGCATCTATCGGATGGGTGATGGAGCGATGTATCCAAAGCATTTACATGCG 153  
 QY 27 AlAGlySerArgThrLeuValValAspLeuValProProGlyArgProAspAla--- 45  
 Db 154 GCGGCGATCGAGACTATCTACGTC-----TGCACAAACCGGACTCTTTC 198  
 QY 46 ----CylLeuValGlyAspValThrAlaProGlyProGluLeuAla-----Ala 60  
 Db 199 GAGGCCCTCGAGAGAAATTCAAAGGAGCGCATATCTCTCCGCAACGGTCAACGCC 258  
 QY 61 AlAlaLeuArgAspAlaLeuValLeuLeuValValHisGluAspValAlaLeuValAla 80  
 Db 259 GTCTCGCGCTTTCACATTCATCTATTTATCTGTC---GAACTGCTGCGCTTCCGCGCT 315  
 QY 81 Val-----AlaProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAsp 96  
 Db 316 GTCTGAGAGAGATGAGACCTTCGACTCGGGTG-----GCGCGGCTTGTAGCTGG 366  
 QY 97 ThrLeuSerValArgThrGlyMetAlaAlaGluLeuAlaAlaHisAlaProGlyValGln 116  
 Db 367 CAGACGAGTGTGAAGCCCGGAGAGGAGGACATTTGAGAGATGTTACCG-----GAG 420  
 QY 117 HisValIGlYLeu-----AsnProMetPhaAlaProAlaAlaGlyMetThrGlyArg 133

Db 421 GATGTGGGATTACAGATGATACATAGTTTCATGGCCGACGATGACGACAGAGGCCGAC 480  
 QY 134 ProValAlaAlaValAlaThrArgAspGlyPro-----GlyValThrAlaLeuLeuArg 151  
 Db 481 CCGCTT--ATTATCATATCATACAGAGGTCCGAAAGAAATGTGAAGATGTGAAGAT 537  
 QY 152 LeuValIGlUGlYGlYGlYArgProValArgLeuThrAlaGluGluHisAspArgThr 171  
 Db 538 GTTTTATAGTCATTCAGAGTGCATATCGTTCACTCTAGCTAGGAGATGATGAAGG 597  
 QY 172 ThrAlaAlaThrGlnAlaLeuThrHisAlaValLeuLeuSerPheGlyLeuAlaLeuAla 191  
 Db 598 ACTGCCAACACCCAGGCTGTGACGATCGCGCTTCTCAGTATGGGTACAGATGCGAG 657  
 QY 192 Arg-LeuGlyValAspValArgAlaLeuAlaAlaThrAlaProProProHisGlnValLe 211  
 Db 658 AATCTTCATCTTACCCATGGAGAAATGACCGCTACGTCTCCGATTC-GAAGTGATCAA 716  
 QY 211 UeuAlaLeuLeuAlaArgValLeuGlyGlySerProGluValTYrGlyAspIleGlnArg 231  
 Db 717 GGTCAACATTCACCTCCGATATCTACTCTGCGCAATGACGATCTACGCCGCTTCCGCT 776  
 QY 231 gSerAsnProArgAlaAlaSer 238  
 Db 777 GCTGAACCCGTCGCCCAATCC 798  
 RESULT 10  
 CF819790  
 LOCUS 859 bp mRNA linear EST 01-APR-2004  
 DEFINITION EST697172 Coccidioides posadasii saprobic phase cDNA library, 2 to 4 kb Coccidioides posadasii cDNA clone CID4071 3' end, mRNA sequence.  
 ACCESSION CF819790  
 VERSION CF819790.1 GI:45925821  
 KEYWORDS EST.  
 SOURCE Coccidioides posadasii  
 ORGANISM Coccidioides posadasii  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Onygenales; Mitosporic Onygenales; Coccidioides.  
 REFERENCE 1 (bases 1 to 859)  
 Gardner, M.J. and Cole, G.T.  
 Analysis of gene expression in Coccidioides posadasii mycelia and spherules via expressed sequence tags  
 JOURNAL Unpublished (2003)  
 COMMENT Other ESTs: EST697173  
 Contact: Gardner MJ  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301 838 3519  
 Fax: 301 838 0208  
 Email: gardner@tigr.org.  
 FEATURES  
 source  
 1. 859  
 Location/Qualifiers  
 /organism="Coccidioides posadasii"  
 /mol\_type="mRNA"  
 /strain="C735"  
 /db\_xref="taxon:199306"  
 /clone="CID4071"  
 /dev\_stage="saprobic phase (mycelia)"  
 /lab\_host="E. coli DH10B, T1 phage resistant"  
 /clone\_lib="Coccidioides posadasii saprobic phase cDNA library, 2 to 4 kb"  
 /note="Vector: pEXpress 1; Site\_1: Not I; Site\_2: Eco RV; Coccidioides posadasii saprobic phase cDNA library, size fractionated cDNA 2 to 4 kb"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.0106 Length: 859  
 Score: 146.00 Matches: 59  
 Percent Similarity: 43.78% Conservative: 29  
 Best Local Similarity: 29.35% Mismatches: 83  
 Query Match: 8.93% Indels: 31



DB: 7 Gaps: 7

US-10-089-514-6 (1-322) x CF819790 (1-859)

OY 5 ProArgerServalValValaIGlyGlySerGlyAlaValaIGlyMetPheAlaGlyLeuLeu 24

Db 268 CCAAGATCCATTGGGATCATCTGGATGGGTGATATACGAGAAGATGATATGCTCAGCGCTTG 327

OY 25 ArgGluAlaGlySerArg-----ThrLeuValVal 34

Db 328 AGCCAAAGGGGGTGGAGACTTAAAAAGACCCCTTAATACCGTGGCTGATGTATTACTACTA 387

OY 35 AspleuValProProProGlyArgProAspAlaCysLeuValGlyAspValThrAlaPro 54

Db 388 GCATTAAATGCGCTGCGACCAACCTTACCATTTATGACACACTGAGGAGAGAAATTCATATCC 447

OY 55 Gly-----ProGluLeuAlaAlaAlaLeuArgAspAlaAspleuVal 68

Db 448 AT-ACAAATATCATATATTATACCAAGGACACTTTGGTCTCCAGAGTTAGGACATCACTT 506

OY 69 LeuLeuAlaValHisGluAspValAlaLeuIysAlaValAlaProValThrArgLeuMet 88

Db 507 ATCTACAGTGTAGAGGACGAGAACCCATCAACAAATTTGGGCGGAGTATAGCCCATCAACC 566

OY 89 ArgProGlyAlaLeuLeuAlaAspThrLeuSerValaGlyThrGlyMetAlaAlaGlyLeu 108

Db 567 AAAGTTGGCGGATGTTGGTGTCAAACTTCCCTGCAAA-----GCCCCAGAGCTG 617

OY 109 AlaAla-----HisAlaProGly--ValGlnHisValGlyLeuAspProwMetPhe 124

Db 618 GCTGTTTGGAGAACCATCTTCTGGGACGTGAGATTAATTTGCGTTCATCTCCATCACT 677

OY 125 AlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValaThrArgAspGlyPro 144

Db 678 GGACCGAAGCGTGAACCTTAAGAGACAGCGTTGCGTTCGATTACACACAGA----- 728

OY 145 GlyValThrAlaLeuLeuArgLeuValGlu-----GlyGlyGlyGlyArgPro 160

Db 729 GCCTCTGACGAGAGCGCTCAGGTTGGTTGTTAAAGTGTGGTCTTTCGAATCCAAATAC 788

OY 161 ValArgLeuThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHis 180

Db 789 GTTATCTTACCGGGGAGAGCATGATGATCAATCACAGCGACACTCAGCGCGTGACGCT 848

OY 181 Ala 181

Db 849 GCT 851

RESULT 11

LOCUS CL959323/c 1407 bp DNA linear GSS 21-SEP-2004

DEFINITION OslFCC002444 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.

ACCESSION CL959323

VERSION CL959323.1 GI:52373397

KEYWORDS GSS.

SOURCE ORGANISM Oryza sativa (indica cultivar-group)

ORIGIN Oryza sativa (indica cultivar-group)

REFERENCE 1 (bases 1 to 1407)

AUTHORS Ma,L., Wang,C.J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M., Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L., Wong,K.K.S., Deng,X.W. and Wang,J.

TITLE An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis

JOURNAL Unpublished (2004)

COMMENT Contact: Chen Chen

Department of Bioinformatic

Beijing Institute of Genomics

Chinese Academy of Sciences, Beijing 101300, China

Tel: 86-10-80481559

Fax: 86-10-80488676  
 Email: chenchen@genomics.org.cn  
 Rice genomic sequence.  
 Class: exon-trapped.  
 Location/Qualifiers  
 1..1407  
 /organism="Oryza sativa (indica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:39946"  
 /clone\_lib="Oryza sativa Expressed Library"  
 /note="Oryza sativa exon trapped genomic sequences"

ORIGIN

Alignment Scores:

Pred. No.:	0.0218	Length:	1407
Score:	145.50	Matches:	97
Percent Similarity:	36.00%	Conservative:	20
Best Local Similarity:	29.85%	Mismatches:	106
Query Match:	8.90%	Indels:	102
DB:	9	Gaps:	15

US-10-089-514-6 (1-322) x CU959323 (1-1407)

```

Oy 17  G|yG|yMetPheaa|aG|yLeuLeuArgG|ua|aG|ySerArgThrLeuVala|aSp|eu 36
Db 1225 GCGCGTGGATTGGT-----CGCCTA 1205

Oy 37  ValProProG|yArgProAspAlaCys|euVal|G|yAspVal|ThrAlaProG|yPro 56
Db 1204 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGATT-----GCTCGTGGCCAGGC--- 1154

Oy 57  Glu|eu|a|a|a|a|e|u|a|a|s|p|a|a|s|p|e|u|a|l|e|u|e|u|a|a|a|H|s|G|u|a|s|p|a| 76
Db 1153 -----GCCGCTGATTAT-----TTG 1139

Oy 77  Al|a|e|u|s|a|l|a|a|a|a|a|p|ro|val|Th|A|r|G|e|u|e|t|a|r|p|ro|G|y|a|l|e|u|e|u|a|a|s|p 96
Db 1138 GCGCAACGCGCGCGCTTCGCGGCTTCATTGCGGTGCGCGCGCG----- 1097

Oy 97  Thr|e|u|s|e|r|a|a|r|g|-----Th|r|G|y|e|t|a| 105
Db 1096 ACTTCGTCCTGACGCTTCGTCGTCGCGCGCGCGCTTCTTGATGTGTGGTGGCATTTGCG 1037

Oy 106 Al|a|G|l|e|u|e|u|a|a|a|H|s|a|a|p|ro|G|y|Val-G|H|s|I|s|a|G|y|e|u|a|n|p|ro|e|t|P|he|a| 125
Db 1036 GCTCGATGCGCGCGCGCGCGCGCGGTTCGCGCGCACTTGGGCGTCTC-----CCCGGTGGC 980

Oy 125 aPro-----Ala|a|a|G|y|e 130
Db 979 TCTCTGGCGCGCGCGCGCTTCGCTTCGCTTCGCTTCGCTTCGCGCGCTGCTGCTGCT 920

Oy 130 tTh|r|G|y|A|r|g|P|ro|Val|a|a|a|Val|a|l|Th|A|r|G|Asp|G|y|P|ro|G|y|Val|Th|r|Ala----- 148
Db 919 CGCTGTGCTGCTGCTGCTGCTGCGCGCGCGCAAGAAAGAACTTGGCGCTGCTGCTGCT 860

Oy 149 -----Leu|e|u|a|r|G|e|u|Val|G|u|-----G|y|G|y|G|y|G|y|A|r|g|P|ro|Val|A|a| 162
Db 859 CCTGCGCTCATCTCTCTGCTGCTTGTGTGAAGAAACAGAGCGGTGGGTGGGTGCTTCG 800

Oy 162 g|L|e|u|Th|r|a|G|l|u|H|s|a|p|a|r|g|Th|r|Th|r|Ala|a|a|Th|r|G|n|a|l|e|u|Th|r|H|s|a|a|a| 182
Db 799 C-----GCGCACACCTCCACGCTTGCCTGCTTCTACGAGAGACTC-----CCCT 755

Oy 182 l|Leu|e|u|s|e|r|P|he|e|y|e|u|a|l|e|u|a|a|a|g|l|e|u|G|y|Val|a|s|p|Val|a|r|g|a|l|e|u|a|a| 202
Db 754 CCTTCTTCCTCATGCGCGGATCGCGCTTGCATGGCGCTGCACCTCCGAGCCTCGCGG 695

Oy 202 aTh|r|Ala|Pro|Pro|H|s|G|Val|e|u|e|u|a|a|e|u|e|u|a|a|e|u|e|u|a|a|r|g|Val|e|u|G|y|s|e 222
Db 694 C-----GCGCTGCGGTCCGCTGCGGCTGCGGCTCTTGAATCTTGG 653

Oy 222 rP|ro|G|u|Val|Tyr|G|y|a|s|p|l|e|G|n|a|r|S|e|r|a|n|----- 233
  
```



Db	652	CCCAATTGGTGGAGAGGGGGCGCGCGCGGGGCGCAACTGATCGCGCGGGCGGGGCGGTCTTCGG	593
Qy	224	-----ProhArg1a1aSer1a1aGh1a1aLeu1aG1	245
Db	592	CGGCGGCGGCGGGGCGCGCGCGGTTCGCGCGCGCTTGAGGGCGCGCGCTTCCTGG	533
Qy	245	u1a1eua1rGserPheal1a1aLeu1aG1Yaspa1pProaPa1rG1a1aPa1aProG1	265
Db	532	CGCC---CGGATCTCTCTCGCCCTTGTAGTGGCAACCAAGATGTCGCGCTCCGTCAGCA	476
Qy	265	YArg1a1aPa1aProG1Yh1a1pProG1Yc1a1aCysa1pG1Yal1aG1Ya1aLeu1a1pG1Ya	285
Db	475	CCTGGCGGAGCGGAGCAACTGCTCGGGGCAACGATGAGCGGTGGCGAGCCGACGATCT	416
Qy	285	1PheG1Yg1Leu1e	289
Db	415	TGACGCTGAACCTC	403

RESULT 12		
AG061086/c		
LOCUS	969 bp	DNA
AG061086		linear
DEFINITION	Pan troglodytes DNA, clone: P1B-049A11.F, genomic survey sequence.	GS5 03-NOV-2001
Accession		

VERSION	AG061086.1	GI:16612888
KEYWORDS	GSS.	
SOURCE	Pan troglodytes (chimpanzee)	
ORGANISM	Pan troglodytes	

REFERENCE  
AUTHORS  
1  
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,

TITLE	BAC end sequences of library PTB
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 969)
AUTHORS	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Tokochi,Y., Watanabe,H. and Sakaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical

**JOURNAL**  
Submitted (02-AUG-2001) Asso Fujitama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suchojo-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: [chimp@gscc.riken.go.jp](mailto:chimp@gscc.riken.go.jp), <http://hsp.gsc.riken.go.jp/>, Tel:81-45-503-9111, Fax:81-45-503-9170)  
**COMMENT**  
Clones are derived from the chimpanzee BAC library PTD This BAC end was generated during the RAD process and may have higher chance of clone tracking errors.  
**PRIMERS**

```

Sequencing: -21M13
LIBRARY
Vector      : PK8145
R.Site 1    : SacI
R.Site 2    : SacI
Location/Qualifiers
1. .969
FEATURES
source

```

```

/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="Caxon:9598"
/clone="PTB-049A11.F"
/sex="male"
/cell_type="lymphoblast"
/clone_id="PTB Chimpanzee Male BAC library"
ORIGIN

```

Pred. No.:	0.0205	969
Score:	143.50	75
Percent Similarity:	36.77%	Conservative: 7
Best local Similarity:	33.63%	Mismatches: 103
Query Match:	8.78%	Indels: 39
DB:	9	Gaps: 9
DS-10-089-514-6 (1-322)	x	AG061086 (1-969)

QY	84	Val1hzhArgMetAcrProG1yAlaLeuLeuAlaAspThrLeuSerValArgThrG1y	103
Db	968	GTGTGTGTGTGTGTGTGGCCGGGG-----CGTTCCTT	936
QY	104	MetAla1aGluMetuAla1aH1aSa1aPProG1yValGlnHis--ValG1yLeuAsnPro	122
Db	935	CCGGCCCGCGCGCCCGCGGTGCGCGCCCGCGCGCTCCACATCGGGGGGACAGCCCC	876
QY	123	MetPhe1aAProAla1aG1yMetThrG1yArgProVal1aAlaVal1aThrArgAsp	142
Db	875	TGCGGCGCTCGCGGCTGGCGGGCGGGCGGGCGCGCTCGTGGCGGCGCGCGCGG	816
QY	143	G1yProG1yVal1aThrAlaLeuLeuArgLeuValGluG1yG1yG1yArg-----Pro	160
Db	815	GGC-----GGCGCGMGCGTTCGCTCGCGGGGGGGGGCGCGCGCGCC	768
QY	161	ValArgLeuThrAlaGluGlnHisAspArgThrThrAla1aThrGlnAlaLeuThrHis	180
Db	767	GTGGCGCCCGCTTC	708
QY	181	AlaValLeuLeuSerPheG1yLeuAlaLeuAlaArgLeuG1yValAspValArgAlaLeu	200
Db	707	GCCGTGGCGCGCGCGCGGTGGGG--GCGGTGGCGCGCTCGCGACCGCGCGCGCGCGCC	650
QY	201	Ala1a1aThr1aP1aP1aP1aP1aP1aGlnValLeuLeuAlaLeuLeuAlaArgValLeuG1y	220
Db	649	CGGGCGCGCGCGCGCGCGCGCGCGCGG-----CGCGGGTGGTGGG	608
QY	221	G1ySerProGluValTyArg1aAsp1aGln1aArgSerAsnProArgAla1aSerAlaArg	240
Db	607	GGG-----CGGGCGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG	575
QY	241	ArgAlaLeuAlaGluAlaLeuAlaArgSerPheAlaAlaLeuValG1yAspAspProAspArg	260
Db	574	GCGCGCGCTTCGCGCGCGCGCGCGCTCGCGCGCGCGCGCTCGCGCGCGCGCGCGCC	515
QY	261	AlaAsp1aP1aProG1yArg-----AlaAspAlaP1aProG1yHisProG1yG1yCyAsp	277
Db	514	CCCGGTGG	455
QY	278	G1yAlaG1yAsnLeuAspG1yValPheG1yG1yLeuArgArgLeuMetG1yProG1yLeu	297
Db	454	GGNGGGGGGGGGCGTGGCGCGGCTC--GGGGCGCGGCGC-----GGCCCCCGCGG	407
QY	298	AlaAlaG1y 300	
Db	406	GAGGGCGGG 398	
RESULT 13			
AY408045/c			
LOCUS			
DEFINITION			
Homo sapiens SUV39H1 gene, VIRTUAL TRANSCRIPT, partial sequence,			
genomic survey sequence.			
ACCESSION			
AY408045			
VERSION			
AY408045.1 GI:39764016			
KEYWORDS			
GSS.			
ORGANISM			
Homo sapiens (human)			
SOURCE			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE			
AUTHORS			
1 (bases 1 to 1086)			
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,			
Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,			
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,			
Adams,M.D. and Cargill,M.			
TITLE			
Inferring nonneutral evolution from human-chimp-mouse orthologous			
gene trios			
JOURNAL			
Science 302 (5652), 1960-1963 (2003)			
PUBMED			
14671302			
REFERENCE			
2 (bases 1 to 1086)			
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,			
Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,			
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,			
Adams,M.D., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,			
AUTHORS			



TITLE Adams, M.D. and Cargill, M.  
 JOURNAL Direct Submission  
 COMMENT Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA  
 This sequence was made by sequencing genomic exons and ordering them based on alignment.  
 FEATURES  
 source Location/Qualifiers  
 1.1086  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 <1..>1086  
 /gene="SOV39H1"  
 /locus\_tag="HGM3093"  
 gene  
 Alignment Scores:  
 Pred. No.: 0.0393 Length: 1086  
 Score: 141.00 Matches: 81  
 Percent Similarity: 44.57% Conservative: 34  
 Best Local Similarity: 31.40% Mismatches: 98  
 Query Match: 8.62% Indels: 45  
 DB: 9 Gaps: 14  
 US-10-089-514-6 (1-322) x AY408045 (1-1086)  
 QY ValValaIGlyGlySerGlyAlaValaGlyGlyMetPheAlaGlyLeuLeuArgGluAlaGly 28  
 DB 926 ATAGTATGGGCGCATCCACGGTGTACACGCTCTCCAGTAGTCCAGGTCAAAGAGTAGGT 867  
 QY 29 SerArgThrLeuValAlaLeuValaLeuValaLeuValaLeuValaLeuValaLeuVala 48  
 DB 866 GCGCGCCCTGACGGTGTGATCTGCGCCCGCCG-----CTCTGCTCTCTC 822  
 QY 49 GlyAspValThrAlaProGlyProGlyLeuAlaAlaLeuArgAspAlaLeuVala 68  
 DB 821 -----TGAGGTATGATCTCTCCACGTAATCATACGAA----- 786  
 QY 69 LeuLeuAlaValaHisGlyAspValaAlaLeuValaAlaValaProValaThrArgLeuMet 88  
 DB 785 -----GCTGTCTTGGCAATCTTCTCCAGGTCGCGACGCCACGCCACGCCCATCATC 732  
 QY 89 Arg-----ProGlyAlaLeuLeuAlaAspThrLeuSerValaArgThrGlyMetAlaAla 106  
 DB 731 CGTGGCGAAGATGACAGAGTCATATCGATACCTTCTGTACACACG---ATTGGGCA 675  
 QY 107 GluLeuAlaAlaHisAlaProGlyValaGlnHisValaGlyLeuAsnProMetPheAlaPro 126  
 DB 674 GTGATAGCCGCGACGCGGAGGAGTTGCACTGTGATGGGACGCGCGCTCGAAAGCG 615  
 QY 127 -----AlaAlaGlyMetThrGly-----ArgProValaAlaAlaValaVala 139  
 DB 614 CACTTGCCCTGTGATGTAGGCAAACTTGTGACGTAGACGCCCGCGGACGACGCTCC 555  
 QY 140 ThrArgAspGlyProGlyValaThrAlaLeuLeu-----ArgLeuValaGluGlyGly 156  
 DB 554 AGTGGGTGCCACAGACAGCTCTGGACCTGCAGCCACAGCAGCCTGGTTGAGGGTAT 495  
 QY 157 GlyGlyArgProValaArgLeuThr-----AlaGluGlnHisAspArgThrThr 172  
 DB 494 GCC-----CTCACCAACACGCTACTGATGATGTAAACGAAAGCCCGCG 450  
 QY 173 AlaAlaThrGlnAlaLeuThrHisAlaValaLeuLeuSerPheGlyLeuAlaLeuAlaArg 192  
 DB 449 AGGGCCGCTCAG-----GTCCACCTCATTTCTTAACAGTATGCGTCCAGATGGCTGG 396  
 QY 193 LeuGlyValaAspValaArgAlaLeuAlaAlaThrAlaProProHisGlnValaLeuLeu 212  
 DB 395 CTGGGATGAGGCTC---CTGCTCCAGCGACGAGCGCGCGCT-----CTGCTT 348  
 QY 213 AlaLeuLeuAlaArgValaLeuGlyGlySer---ProGlyValaGlyArgPheGlnArg 231  
 DB 347 GGCCTTCTGACACGAGTATGGCCAAAGCTTGGGTGACAGGTGCGGGGGGTCTTTGACCG 288

QY 231 gSerAsnProArgAlaAlaSerAlaArgAlaLeuAlaGluAlaLeuArg 248  
 DB 287 GTGTGTCGCCCGGAGAGAGCTCCCTTTCTTAAGTCCTTGTGAGACTGCTTAGG 236  
 RESULT 14  
 CR595067/c 1684 bp mRNA linear HTC 21-JUL-2004  
 LOCUS CR595067/c  
 DEFINITION full-length cDNA clone CS0DL012Y112 of B cells (Ramos cell line)  
 Cot 25-normalized of Homo sapiens (human).  
 ACCESSION CR595067  
 VERSION CR595067.1 GI:50475874  
 KEYWORDS HTC; CNSLT\_cDNA.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 1684)  
 Li, W.B., Gruber, C., Jesse, J., and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished  
 Contact : Feng Liang Email : fliang@liffech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue 2 (bases 1 to 1684)  
 REFERENCE  
 AUTHORS Genoscope.  
 TITLES Direct Submission  
 JOURNAL Submitted (20-JUN-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 COMMENT  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of invitrogen.  
 FEATURES  
 source Location/Qualifiers  
 1.1684  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DL012Y112"  
 /tissue\_type="B cells (Ramos cell line) Cot 25-normalized"  
 /plasmid="pCMVSPORT\_6"  
 Alignment Scores:  
 Pred. No.: 0.0683 Length: 1684  
 Score: 141.00 Matches: 81  
 Percent Similarity: 44.57% Conservative: 34  
 Best Local Similarity: 31.40% Mismatches: 98  
 Query Match: 8.62% Indels: 45  
 DB: 3 Gaps: 14  
 US-10-089-514-6 (1-322) x CR595067 (1-1684)  
 QY 9 ValValaIGlyGlySerGlyAlaValaGlyGlyMetPheAlaGlyLeuLeuArgGluAlaGly 28  
 DB 928 ATAGTATGGGCGCATCCACGGTGTACACGCTCTCCAGTAGTCCAGGTCAAAGAGTAGGT 869  
 QY 29 SerArgThrLeuValAlaLeuValaLeuValaLeuValaLeuValaLeuValaLeuVala 48  
 DB 868 GCGCGCCCTGACGGTGTGATCTGCGCCCGCCG-----CTCTGCTCTCTC 824  
 QY 49 GlyAspValThrAlaProGlyProGlyLeuAlaAlaAlaLeuArgAspAlaLeuVala 68  
 DB 823 -----TGAGGTATGATCTCTCCACGTAATCATACGAA----- 788  
 QY 69 LeuLeuAlaValaHisGlyAspValaAlaLeuValaAlaValaProValaThrArgLeuMet 88  
 DB 787 -----GCTGTCTTGGCAATCTTCTCCAGGTCGCGACGCCACGCCCATCATC 734  
 QY 89 Arg-----ProGlyAlaLeuLeuAlaAspThrLeuSerValaArgThrGlyMetAlaAla 106







